SEQ ID 8570 (GBS271) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 51 (lane 8; MW 31.3kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 55 (lane 6; MW 56.3kDa) and in Figure 62 (lane 10; MW 56.3kDa).

GBS271-GST was purified as shown in Figure 210, lane 8.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 396

A DNA sequence (GBSx0430) was identified in *S.agalactiae* <SEQ ID 1287> which encodes the amino acid sequence <SEQ ID 1288>. Analysis of this protein sequence reveals the following:

```
Possible site: 22
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -6.74   Transmembrane    9 -  25 (   5 -  28)
   INTEGRAL    Likelihood = -5.84   Transmembrane   97 - 113 (  92 - 122)
   INTEGRAL    Likelihood = -5.47   Transmembrane   37 -  53 (  35 -  61)
   INTEGRAL    Likelihood = -2.55   Transmembrane  220 - 236 ( 220 - 238)
   INTEGRAL    Likelihood = -1.65   Transmembrane   64 -  80 (  63 -  81)
   INTEGRAL    Likelihood = -1.28   Transmembrane  193 - 209 ( 192 - 209)
   INTEGRAL    Likelihood = -0.53   Transmembrane  125 - 141 ( 125 - 141)

----- Final Results -----
               bacterial membrane --- Certainty=0.3697(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAC73593 GB:AE000155 putative metal resistance protein
          [Escherichia coli K12]
 Identities = 128/252 (50%), Positives = 186/252 (73%)

Query: 5   NSISLMSLLMASSLVLITLFFSYWQKLNLEKEVIISAIRAVIQLLAVGFLLDYIFGYQNP 64
           ++I+  SL +A  LV++ +  S+ +KL LEK+++ S  RA+IQL+ VG++L YIF   +
Sbjct: 13  HNITNESLALALMLVVVAILISHKEKLALEKDILWSVGRAIIQLIIVGYVLKYIFSVDDA 72

Query: 65  IFTALLMLFMIINASYNAAKRGKGINKGFVISFIAIGSGTIITLSVLIFSGILKFVPNQM 124
             T L++LF+  NA++NA KR K I K F+ SFIAI  G  ITL+VLI SG ++F+P Q+
Sbjct: 73  SLTLLMVLFICFNAAWNAQKRSKYIAKAFISSFIAITVGAGITLAVLILSGSIEFIPMQV 132

Query: 125 IPVGGMIISNSMVAIGLCYKQLLSEFRSKQEEVETKLALGADILPASIDIIRDVIKTGMV 184
           IP+ GMI  N+MVA+GLCY  L      S+Q++++ KL+LGA    AS  +IRD I+  ++
Sbjct: 133 IPIAGMIAGNAMVAVGLCYNNLGQRVISEQQQIQEKLSLGATPKQASAILIRDSIRAALI 192

Query: 185 PTIDSAKTLGIVSLPGMMTGLILAGTSPIQAVKYQMMVTFMLLATTSIASFVATYLAYKI 244
           PT+DSAKT+G+VSLPGMM+GLI AG  P++A+KYQ+MVTFMLL+T S+++ +A YL Y+
Sbjct: 193 PTVDSAKTVGLVSLPGMMSGLIFAGIDPVKAIKYQIMVTFMLLSTASLSTIIACYLTYRK 252

Query: 245 FFNNRKQLVVTK 256
           F+N+R QLVVT+
Sbjct: 253 FYNSRHQLVVTQ 264
```

No corresponding DNA sequence was identified in *S.pyogenes.*

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 397

A DNA sequence (GBSx0431) was identified in *S.agalactiae* <SEQ ID 1289> which encodes the amino acid sequence <SEQ ID 1290>. This protein is predicted to be SUGAR TRANSPORT ATP-BINDING PROTEIN. (b0490). Analysis of this protein sequence reveals the following:

```
Possible site: 56
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1903(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAC73592 GB:AE000155 putative ATP-binding component of a
           transport system [Escherichia coli K12]
 Identities = 95/202 (47%), Positives = 142/202 (70%), Gaps = 2/202 (0%)

Query: 4   LTFKHVDFKTDDKLVLNDINFAIDEGDFVSIVGPSGSGKSTVLKLASGLMSPTAGHIFFD 63
           L  ++V +   D  +LN+INF++  G+F  I GPSG GKST+LK+ + L+SPT+G + F+
Sbjct: 8   LQLQNVGYLAGDAKILNNINFSLRAGEFKLITGPSGCGKSTLLKIVASLISPTSGTLLFE 67

Query: 64  GKDLNQLEPIESRKMISYCFQTPHLFGNTVEDNISFPYHIRHEKVDYRRVDDLFQRFEMD 123
           G+D++ L+P   R+ +SYC QTP LFG+TV DN+ FP+ IR+ + D     D  +RF +
Sbjct: 68  GEDVSTLKPEIYRQQVSYCAQTPTLFGDTVYDNLIFPWQIRNRQPDPAIFLDFLERFALP 127

Query: 124 QSYLKQDVKKLSGGEKQRIALIRQLLFEPKVLLLDEVTSALDNHNKAIVEKVI-KSLHDK 182
            S L +++ +LSGGEKQRI+LIR L F PKVLLLDE+TSALD  NK  V ++I + + ++
Sbjct: 128 DSILTKNIAELSGGEKQRISLIRNLQFMPKVLLLDEITSALDESNKHNVNEMIHRYVREQ 187

Query: 183 GITILWITHDEEQSRRFANKVL 204
            I +LW+THD+++    A+KV+
Sbjct: 188 NIAVLWVTHDKDEINH-ADKVI 208
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1291> which encodes the amino acid sequence <SEQ ID 1292>. Analysis of this protein sequence reveals the following:

```
Possible site: 61
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2053(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 73/214 (34%), Positives = 133/214 (62%), Gaps = 9/214 (4%)

Query: 4   LTFKHVD--FKTDDKLVLNDINFAIDEGDFVSIVGPSGSGKSTVLKLASGLMSPTAGHIF 61
           +TF +V   F+      VL +INF ++EG F +++G SGSGKST+L + +GL+  ++G I+
Sbjct: 6   ITFNNVSKTFEDSGTQVLKNINFDLEEGKFYTLLGASGSGKSTILNIMAGLLDASSGDIY 65

Query: 62  FDGKDLNQLEPIESRKMISYCFQTPHLFGN-TVEDNISFPYHIR--HEKVDYRRVDDLFQ 118
            DG+ +N L PI  R  I   FQ   LF + TV +N++F   ++   +K   +RV +  +
Sbjct: 66  LDGERINDL-PINKRD-IHTVFQNYALFPHMTVFENVAFALKLKKVDKKEIAKRVKETLK 123

Query: 119 RFEMDQSYLKQDVKKLSGGEKQRIALIRQLLFEPKVLLLDEVTSALDNHNKAIVEKVIKS 178
             ++ + +  + ++KLSGG++QR+A+ R ++ +P+V+LLDE  SALD   +  ++  ++
Sbjct: 124 MVQL-EGFENRSIQKLSGGQRQRVAIARAIINQPRVVLLDEPLSALDLKLRTEMQYELRE 182

Query: 179 LHDK-GITILWITHDEEQSRRFANKVLKVVNGSI 211
           L  + GIT +++THD+E++   ++ +  +  G I
Sbjct: 183 LQQRLGITFVFVTHDQEEALAMSDWIFVMNEGEI 216
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 398

A DNA sequence (GBSx0432) was identified in *S.agalactiae* <SEQ ID 1293> which encodes the amino acid sequence <SEQ ID 1294>. Analysis of this protein sequence reveals the following:

```
Possible site: 32
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0658(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 399

A DNA sequence (GBSx0434) was identified in *S.agalactiae* <SEQ ID 1295> which encodes the amino acid sequence <SEQ ID 1296>. This protein is predicted to be deda protein (dedA). Analysis of this protein sequence reveals the following:

```
Possible site: 58
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-12.05   Transmembrane  186 - 202 ( 178 - 208)
   INTEGRAL    Likelihood = -8.81   Transmembrane   65 -  81 (  61 -  89)
   INTEGRAL    Likelihood = -7.54   Transmembrane   26 -  42 (  24 -  47)
   INTEGRAL    Likelihood = -0.37   Transmembrane  152 - 168 ( 152 - 168)

----- Final Results -----
               bacterial membrane --- Certainty=0.5819(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAC75377 GB:AE000320 orf, hypothetical protein [Escherichia coli K12]
 Identities = 91/211 (43%), Positives = 131/211 (61%), Gaps = 7/211 (3%)

Query: 2   FLIDFILHIDTHIYAMANTVGNWTYLLLFLVIFVETGAVIFPFLPGDSLLFAAGALAANP 61
           FLIDFILHID H+  +    G W Y +LFL++F ETG V+ PFLPGDSLLF AGALA+
Sbjct: 6   FLIDFILHIDVHLAELVAEYGVWVYAILFLILFCETGLVVTPFLPGDSLLFVAGALASLE 65

Query: 62  KMSFNIVTFLIIFFIAAFIGDSCNFLIGRTFGYRFIKHP---FFRRFIKEKNIRDAELYF 118
               N+   +++  IAA +GD+ N+ IGR FG +   +P    FRR   +K       ++
Sbjct: 66  TNDLNVHMMVVLMLIAAIVGDAVNYTIGRLFGEKLFSNPNSKIFRRSYLDK----THQFY 121

Query: 119 EKKGTAAIILGRYIPIIRTFVPFVAGISQLPPKVFIKRAFIAALSWSVIATGSGFLFGNI 178
           EK G   IIL R++PI+RTF PFVAG+  +  + F     I AL W ++ T +G+ FG I
Sbjct: 122 EKHGGKTIILARFVPIVRTFAPFVAGMGHMSYRHFAAYNVIGALLWVLLFTYAGYFFGTI 181

Query: 179 PFVKQHFSLIILGIVFVTLIPVLISGVKSYR 209
           P V+ +  L+I+GI+ V+++P +I  ++  R
Sbjct: 182 PMVQDNLKLLIVGIIVVSILPGVIEIIRHKR 212
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 400

A DNA sequence (GBSx0435) was identified in *S.agalactiae* <SEQ ID 1297> which encodes the amino acid sequence <SEQ ID 1298>. Analysis of this protein sequence reveals the following:

```
Possible site: 18
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3100(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 401

A DNA sequence (GBSx0436) was identified in *S.agalactiae* <SEQ ID 1299> which encodes the amino acid sequence <SEQ ID 1300>. This protein is predicted to be DNA-entry nuclease. Analysis of this protein sequence reveals the following:

```
Possible site: 51
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3990(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9323> which encodes amino acid sequence <SEQ ID 9324> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAA38134 GB:X54225 membrane nuclease [Streptococcus pneumoniae]
 Identities = 87/157 (55%), Positives = 110/157 (69%), Gaps = 1/157 (0%)

Query: 1   MLDRTIRQYQNRRDTTLPDANWKPLGWHQVAT-NDHYGHAVDKGHLIAYALAGNFKGWDA 59
           +L +  RQY+NR++T     +W P GWHQV      Y HAVD+GHL+ YAL G   G+DA
Sbjct: 116 LLSKATRQYKNRKETGNGSTSWTPPGWHQVKNLKGSYTHAVDRGHLLGYALIGGLDGFDA 175

Query: 60  SVSNPQNVVTQTAHSNQSNQKINRGQNYYESLVRKAVDQNKRVRYRVTPLYRNDTDLVPF 119
           S SNP+N+  QTA +NQ+  + + GQNYYES VRKA+DQNKRVRYRVT  Y ++ DLVP
Sbjct: 176 STSNPKNIAVQTAWANQAQAEYSTGQNYYESKVRKALDQNKRVRYRVTLYYASNEDLVPS 235

Query: 120 AMHLEAKSQDGTLEFNVAIPNTQASYTMDYATGEITL 156
           A  +EAKS DG LEFNV +PN Q    +DY TGE+T+
Sbjct: 236 ASQIEAKSSDGELEFNVLVPNVQKGLQLDYRTGEVTV 272
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1301> which encodes the amino acid sequence <SEQ ID 1302>. Analysis of this protein sequence reveals the following:

```
     Possible site: 42
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
               bacterial outside --- Certainty=0.3000(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:CAA38134 GB:X54225 membrane nuclease [Streptococcus pneumoniae]
 Identities = 89/135 (65%), Positives = 104/135 (76%), Gaps = 1/135 (0%)

Query: 25  SPAGWHRLHHLKGSYDHAVDRGHLLGYALVGGLKGFDASTGNPDNIATQLSWANQANKPY 84
            +P GWH++ +LKGSY HAVDRGHLLGYAL+GGL GFDAST NP NIA Q +WANQA   Y
Sbjct: 138 TPPGWHQVKNLKGSYTHAVDRGHLLGYALIGGLDGFDASTSNPKNIAVQTAWANQAQAEY 197

Query: 85  LTGQNYYEGLVRRALDKGHRVRYRVTLLY-DGDNLLASGSHLEAKSSDDSLTFNVFVPNV 143
             TGQNYYE  VR+ALD+  RVRYRVTL Y   ++L+ S S +EAKSSD  L FNV VPNV
Sbjct: 198 STGQNYYESKVRKALDQNKRVRYRVTLYYASNEDLVPSASQIEAKSSDGELEFNVLVPNV 257

Query: 144 QAGLTADYRTGQIAI 158
            Q GL  DYRTG++ +
Sbjct: 258 QKGLQLDYRTGEVTV 272
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 73/135 (54%), Positives = 92/135 (68%), Gaps = 2/135 (1%)

Query: 24  PLGWHQVA-TNDHYGHAVDKGHLIAYALAGNFKGWDASVSNPQNVVTQTAHSNQSNQKIN 82
            P GWH++      Y HAVD+GHL+ YAL G  KG+DAS  NP N+ TQ + +NQ+N+
Sbjct: 26  PAGWHRLHHLKGSYDHAVDRGHLLGYALVGGLKGFDASTGNPDNIATQLSWANQANKPYL 85

Query: 83  RGQNYYESLVRKAVDQNKRVRYRVTPLYRNDTDLVPFAMHLEAKSQDGTLEFNVAIPNTQ 142
             GQNYYE LVR+A+D+  RVRYRVT LY  D +L+    HLEAKS D +L FNV +PN Q
Sbjct: 86  TGQNYYEGLVRRALDKGHRVRYRVTLLYDGD-NLLASGSHLEAKSSDDSLTFNVFVPNVQ 144

Query: 143 ASYTMDYATGEITLN 157
            A  T DY TG+I +N
Sbjct: 145 AGLTADYRTGQIAIN 159
```

SEQ ID 9324 (GBS656) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 186 (lane 10; MW 57kDa).

GBS656-GST was purified as shown in Figure 236, lane 4.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 402

A DNA sequence (GBSx0437) was identified in *S.agalactiae* <SEQ ID 1303> which encodes the amino acid sequence <SEQ ID 1304>. Analysis of this protein sequence reveals the following:

```
Possible site: 13
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
               bacterial outside --- Certainty=0.3000(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9321> which encodes amino acid sequence <SEQ ID 9322> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1305> which encodes the amino acid sequence <SEQ ID 1306>. Analysis of this protein sequence reveals the following:

```
Possible site: 47
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.5350(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 24/73 (32%), Positives = 37/73 (49%), Gaps = 2/73 (2%)

Query: 1   MFYMKLANRLSLAATIVNEANANSPFGIIIHSDKAENVEWNDFETQFPDLFNSPKKEESP 60
           + YMKLA    L  TI+ E +  SPF  I+H+D A N++    E       N    +++P
Sbjct: 80  ILYMKLAKENHLPVTIIITETHMTSPFAFILHTDHAINLKETRLEVILKQTKNDQLSKQTP 139

Query: 61  K--KSLWQHFFSQ 71
           +  KS W+ F  +
Sbjct: 140 EKTKSFWKRFLKK 152
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 403

A DNA sequence (GBSx0438) was identified in *S.agalactiae* <SEQ ID 1307> which encodes the amino acid sequence <SEQ ID 1308>. This protein is predicted to be Isopentenyl-diphosphate delta-isomerase. Analysis of this protein sequence reveals the following:

```
Possible site: 39
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1649(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAG20030 GB:AE005083 isopentenyl pyrophosphate isomerase; Idi
           [Halobacterium sp. NRC-1]
 Identities = 24/77 (31%), Positives = 40/77 (51%)

Query: 14  TGLTLNRDQNIPQGLFHLVVDVILFHEDGDVLMMKRHPKKKAFPAYFEATAGGSALKGEN 73
           TGL    D +   G+ H     +LF EDG VL+ +R  +K+ +  +++ T     ++G++
Sbjct: 42  TGLANRLDAHTGDGVRHRAFTCLLFDEDGRVLLAQRADRKRLWDTHWDGTVASHPIEGQS 101

Query: 74  AKQAILRELKEETGIVP 90
              A  + L EE GI P
Sbjct: 102 QVDATRQRLAEELGIEP 118
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 404

A DNA sequence (GBSx0439) was identified in *S.agalactiae* <SEQ ID 1309> which encodes the amino acid sequence <SEQ ID 1310>. This protein is predicted to be phosphoserine phosphatase (serB). Analysis of this protein sequence reveals the following:

```
Possible site: 35
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0613(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB50876 GB:AL096844 putative phosphoserine phosphatase
           [Streptomyces coelicolor A3(2)]
 Identities = 96/193 (49%), Positives = 132/193 (67%)

Query: 5   LLVMDVDSTLIMEEAIDLLAIEAGVGKQVAALTDAAMRGELDFEEALKKRVALLKGLPVT 64
           L+VMDVDSTLI +E I+L A  AG   +VA +T AAMRGELDFE++L  RVALL GL  +
Sbjct: 183 LVVMDVDSTLIQDEVIELFAAHAGCEDEVAEVTAAAMRGELDFEQSLHARVALLAGLDAS 242

Query: 65  ILTDILSSIHFTPGAYELIKECHKRQMKVGLVSGGFHETIDILAKQLQVDYVKANRLGVK 124
           ++  + + +  TPGA  LI+   +   +VG+VSGGF +  D L +QL +D+ +AN L +
Sbjct: 243 VVDKVRAEVRLTPGARTLIRTLKRLGYQVGVVSGGFTQVTDALQEQLGLDFAQANTLEIV 302

Query: 125 GGFLTGEVEGEIVTKEVKKIKLKEWASENHLDLSQTIAMGDGANDLPMIKSAGVGIAFCA 184
            G LTG V GEIV +  K   L+ +A+   + LSQT+A+GDGANDL M+ +AG+G+AF A
Sbjct: 303 DGRLTGRVTGEIVDRAGKARLLRRFAAAAGVPLSQTVAIGDGANDLDMLNAAGLGVAFNA 362

Query: 185 KPIVREEAAYQIN 197
           KP+VRE A   +N
Sbjct: 363 KPVVREAAHTAVN 375
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 405

A DNA sequence (GBSx0440) was identified in *S.agalactiae* <SEQ ID 1311> which encodes the amino acid sequence <SEQ ID 1312>. Analysis of this protein sequence reveals the following:

```
Possible site: 23
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-17.88   Transmembrane    5 -  21 (   1 -  29)

----- Final Results -----
               bacterial membrane --- Certainty=0.8153(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAB06924 GB:AP001518 unknown conserved protein [Bacillus halodurans]
 Identities = 122/553 (22%), Positives = 265/553 (47%), Gaps = 12/553 (2%)
```

```
Query: 7   LLLVAIVLLVIIAYVVGVVIRKRNDTLIANLETRKQELVDLPVQEEIEQVKLLHLIGQSQ 66
           +++ ++++L +  +V G + RK     +  LE  K +++  P+ +EI +VK L + G+++
Sbjct: 3   IVVFSLLVLTVTFFVYGALRRKAFYKRVDKLEDWKNDILQRPIPDEIGKVKGLTMSGETE 62

Query: 67  STFREWNQKWTDLSTNSFKDIDFHLVEAENLNDSFNFVRAKHEIDNVDSQLTIIEEDIVS 126
             F  W   W D+      +++  L + E+  + + F +AK  +D ++ +L  IEE +
Sbjct: 63  EKFEVWRSDWDDIVGVILPNVEEQLFDVEDFANKYRFQKAKALLDTIEQRLHSIEEQLKI 122

Query: 127 IREALEVLKEQEEKNSARVTHALDLYETLQKSISEKEDNYGTTMPEIEKQLKNIEAEFSH 186
           + + ++VL + EE+N   +    +L + L K    +  +  ++    +++L+
Sbjct: 123 MVDDIQVLVQSEEQNRTEIGSVRELQQKLIKEAITRRGSLSSSAKVFDEKLEKANELLQA 182

Query: 187 FVTLNSTGDPIEASEVLNKAEEHTIALGQITEQIPAIVAKLEDDFPDQLDDLETGYRRLL 246
           F      G+ I+ASEVL +A+E    +  + + +P +  +L+ + P +L +L+ G R +
Sbjct: 183 FDERTEKGNYIQASEVLEEAKELLGQIEHLLKIVPGLFVELQTNIPAELTNLKNGLRDME 242

Query: 247 EENYHFPEKDIEQRFQEVREAIRSNSDGLVSLDLDRARDENEHIQEKIDKLYDIFEREIA 306
           E  +      I+ + + + E      + L  L+ +   +E   I+E +++++++ E+E+
Sbjct: 243 EAGFFLETFAIDSQMERLEEKRVELLEQLTVLECNGMEEEINFIEESMEQMFELLEKEVE 302

Query: 307 AYKVAHKDSKIIPQFLAHAKSNNEQLGH---EIKRLSAKYILNENESLSLRSFTNDLEEI 363
           A    ++ + ++P          E+L H   E + +   Y L E E +  +    +L+E+
Sbjct: 303 A---KNEITILLPNLREDLTKTEEKLTHLKEETESVQLSYRLAEEELVFQQKLGKELKEL 359

Query: 364 ETKVLPSVENFGQEASPYTHLQILFERTLKTLTTVEENQMEVFEAVKTIESVETRARQNM 423
             ++    E   ++   ++ ++ + E   + LT  +    +  E++ ++   E +A++ +
Sbjct: 360 RQQLQVIDEVTEEQKQTFSSVRSMLEEWREGLTACQNKIEQAQESLNSLRKDELKAKEEL 419

Query: 424 DKYVNKLHMIKRFMEKRNLPGIPQDFLSTFFTTSSQIEALINELSRGRIDIEAVSRLNDV 483
            +   KL   KR ++K N+PG+P+  L        ++   I +LS   +++  V+ L D
Sbjct: 420 KQLKEKLLEDKRLVQKSNIPGLPETLLHRLEDGEQKLAQAIAKLSDVPLEMGRVTALVDE 479

Query: 484 TTNAIANLEQATYLVVQDATLTEQLLQYSNRYRSFEQNVQKSFEQALYLFEVEHNYKASF 543
               I       +  ++ A L E ++QY NRYRS    V+K    A  LF         +
Sbjct: 480 AQGLIHENSSILHETIEKARLAEHVIQYGNRYRSRSAEVKKRLSNAEELFRA-----FEY 534

Query: 544 DE-ISYALETVEP 555
           DE I  A++ +EP
Sbjct: 535 DEAIEMAVQAIEP 547
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1313> which encodes the amino acid sequence <SEQ ID 1314>. Analysis of this protein sequence reveals the following:

```
     Possible site: 23
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-18.04    Transmembrane     5 -   21 (   1 -   29)

----- Final Results -----
               bacterial membrane --- Certainty=0.8217(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:BAB06924 GB:AP001518 unknown conserved protein [Bacillus halodurans]
 Identities = 131/555 (23%), Positives = 269/555 (47%), Gaps = 16/555 (2%)

Query: 7   LLIVAIVLLVIIAYLVGVIIRKRNDSLITSLEERKQALFALPVNDEIEEVKSLHLIGQSQ 66
           +++ ++++L +  ++ G + RK     +  LE+ K  +   P+ DEI +VK L + G+++
Sbjct: 3   IVVFSLLVLTVTFFVYGALRRKAFYKRVDKLEDWKNDILQRPIPDEIGKVKGLTMSGETE 62

Query: 67  TSFREWNQKWVDLTVNSFADIENHIFEAENLNDTFNFIRAKHEINSVESQLNLVEEDIAS 126
             F  W   W D+      ++E  +F+ E+  + + F +AK  ++++E +L+ +EE +
Sbjct: 63  EKFEVWRSDWDDIVGVILPNVEEQLFDVEDFANKYRFQKAKALLDTIEQRLHSIEEQLKI 122

Query: 127 IREALNILKEQEEKNSARVTHALDLYEKLQASISENEDNFGSTMPEIDKQMKNIETEFSQ 186
           + + + +L + EE+N   +    +L +KL         +  S+    D++++
```

```
Sbjct: 123 MVDDIQVLVQSEEQNRTEIGSVRELQQKLIKEAITRRGSLSSSAKVFDEKLEKANELLQA 182

Query: 187 FVALNSSGDPVEASEVLDRAEEHTIALGQITEQIPAIVAKLEDDFPDQLDDLETGYRRLL 246
           F      G+ ++ASEVL+ A+E    +  + + +P +  +L+ + P +L +L+ G R +
Sbjct: 183 FDERTEKGNYIQASEVLEEAKELLGQIEHLLKIVPGLFVELQTNIPAELTNLKNGLRDME 242

Query: 247 EENYHFPEKNIEARFQEIRESIRANSSELVTLDLDRAREENTHIQERIDSLYEVFEREIA 306
           E  +      I+++ + + E       +L  L+ +   EE   I+E ++ ++E+ E+E
Sbjct: 243 EAGFFLETFAIDSQMERLEEKRVELLEQLTVLECNGMEEEINFIEESMEQMFELLEKE-- 300

Query: 307 AYKVAAKN--SKMLPRYLEHVKRNNEQ---LKDEIARLSRKYILSETESLTVKAFEKDIK 361
              V AKN  + +LP   E + +  E+   LK+E   +   Y L+E E +  +   K++K
Sbjct: 301 ---VEAKNEITILLPNLREDLTKTEEKLTHLKEETESVQLSYRLAEEELVFQQKLGKELK 357

Query: 362 EIEDSTLAVAEQFGLQEKPFSELQVTFERSIKTLTNVESGQMDVFAAVKDIEKIESQARH 421
           E+      + E    Q++ FS ++   E   + LT  ++       ++  + K E +A+
Sbjct: 358 ELRQQLQVIDEVTEEQKQTFSSVRSMLEEWREGLTACQNKIEQAQESLNSLRKDELKAKE 417

Query: 422 NLDVYVTQLHMIKRYMEKRHLPGIPQDFLSAFFTTSSQLEALMDELSRGRINIEAVSRLS 481
            L     +L   KR ++K ++PG+P+  L        +L   + +LS   + +  V+ L
Sbjct: 418 ELKQLKEKLLEDKRLVQKSNIPGLPETLLHRLEDGEQKLAQAIAKLSDVPLEMGRVTALV 477

Query: 482 EVATVAIANLEDLTYQVVQNATLTEQLLQYSNRYRSFEAGVQSSFEHALRLFEVENDYQA 541
           + A   I     + ++ ++ A L E ++QY NRYRS  A V+    +A  LF
Sbjct: 478 DEAQGLIHENSSILHETIEKARLAEHVIQYGNRYRSRSAEVKKRLSNAEELFRA-----F 532

Query: 542 SFDE-ISYALETVEP 555
            +DE I  A++ +EP
Sbjct: 533 EYDEAIEMAVQAIEP 547
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 429/574 (74%), Positives = 503/574 (86%)

Query: 1   MSSGIILLLVAIVLLVIIAYVVGVVIRKRNDTLIANLETRKQELVDLPVQEEIEQVKLLH 60
           MSSGIILL+VAIVLLVIIAY+VGV+IRKRND+LI +LE RKQ L  LPV +EIE+VK LH
Sbjct: 1   MSSGIILLIVAIVLLVIIAYLVGVIIRKRNDSLITSLEERKQALFALPVNDEIEEVKSLH 60

Query: 61  LIGQSQSTFREWNQKWTDLSTNSFKDIDFHLVEAENLNDSFNFVRAKHEIDNVDSQLTII 120
           LIGQSQ++FREWNQKW DL+ NSF DI+ H+ EAENLND+FNF+RAKHEI++V+SQL ++
Sbjct: 61  LIGQSQTSFREWNQKWVDLTVNSFADIENHIFEAENLNDTFNFIRAKHEINSVESQLNLV 120

Query: 121 EEDIVSIREALEVLKEQEEKNSARVTHALDLYETLQKSISEKEDNYGTTMPEIEKQLKNI 180
           EEDI SIREAL +LKEQEEKNSARVTHALDLYE LQ SISE EDN+G+TMPEI+KQ+KNI
Sbjct: 121 EEDIASIREALNILKEQEEKNSARVTHALDLYEKLQASISENEDNFGSTMPEIDKQMKNI 180

Query: 181 EAEFSHFVTLNSTGDPIEASEVLNKAEEHTIALGQITEQIPAIVAKLEDDFPDQLDDLET 240
           E EFS FV LNS+GDP+EASEVL++AEEHTIALGQITEQIPAIVAKLEDDFPDQLDDLET
Sbjct: 181 ETEFSQFVALNSSGDPVEASEVLDRAEEHTIALGQITEQIPAIVAKLEDDFPDQLDDLET 240

Query: 241 GYRRLLEENYHFPEKDIEQRFQEVREAIRSNSDGLVSLDLDRARDENEHIQEKIDKLYDI 300
           GYRRLLEENYHFPEK+IE RFQE+RE+IR+NS  LV+LDLDRAR+EN HIQE+ID LY++
Sbjct: 241 GYRRLLEENYHFPEKNIEARFQEIRESIRANSSELVTLDLDRAREENTHIQERIDSLYEV 300

Query: 301 FEREIAAYKVAHKDSKIIPQFLAHAKSNNEQLGHEIKRLSAKYILNENESLSLRSFTNDL 360
           FEREIAAYKVA K+SK++P++L H K NNEQL  EI RLS KYIL+E ESL++++F  D+
Sbjct: 301 FEREIAAYKVAAKNSKMLPRYLEHVKRNNEQLKDEIARLSRKYILSETESLTVKAFEKDI 360

Query: 361 EEIETKVLPSVENFGQEASPYTHLQILFERTLKTLTTVEENQMEVFEAVKTIESVETRAR 420
           +EIE   L   E FG +  P++ LQ+ FER++KTLT VE  QM+VF AVK IE +E++AR
Sbjct: 361 KEIEDSTLAVAEQFGLQEKPFSELQVTFERSIKTLTNVESGQMDVFAAVKDIEKIESQAR 420

Query: 421 QNMDKYVNKLHMIKRFMEKRNLPGIPQDFLSTFFTTSSQIEALINELSRGRIDIEAVSRL 480
            N+D YV +LHMIKR+MEKR+LPGIPQDFLS FFTTSSQ+EAL++ELSRGRI+IEAVSRL
Sbjct: 421 HNLDVYVTQLHMIKRYMEKRHLPGIPQDFLSAFFTTSSQLEALMDELSRGRINIEAVSRL 480

Query: 481 NDVTTNAIANLEQATYLVVQDATLTEQLLQYSNRYRSFEQNVQKSFEQALYLFEVEHNYK 540
           ++V T AIANLE  TY VVQ+ATLTEQLLQYSNRYRSFE  VQ SFE AL LFEVE++Y+
```

```
Sbjct: 481 SEVATVAIANLEDLTYQVVQNATLTEQLLQYSNRYRSFEAGVQSSFEHALRLFEVENDYQ 540

Query: 541 ASFDEISYALETVEPGVTDRFVTSYEKTQERIRF 574
           ASFDEISYALETVEPGVTDRFV SYEKT+E IRF
Sbjct: 541 ASFDEISYALETVEPGVTDRFVNSYEKTREHIRF 574
```

SEQ ID 1312 (GBS642) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 142 (lane 2-4; MW 27kDa).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 406

A DNA sequence (GBSx0441) was identified in *S.agalactiae* <SEQ ID 1315> which encodes the amino acid sequence <SEQ ID 1316>. Analysis of this protein sequence reveals the following:

```
Possible site: 41
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.2471(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9671> which encodes amino acid sequence <SEQ ID 9672> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAA91553 GB:Z67740 DNA gyrase [Streptococcus pneumoniae]
 Identities = 574/650 (88%), Positives = 618/650 (94%), Gaps = 2/650 (0%)

Query: 1   MTEETKNMEQRAQEYDASQIQVLEGLEAVRMRPGMYIGSTSKEGLHHLVWEIVDNSIDEA 60
           MTEE KN++  AQ+YDASQIQVLEGLEAVRMRPGMYIGSTSKEGLHHLVWEIVDNSIDEA
Sbjct: 1   MTEEIKNLQ--AQDYDASQIQVLEGLEAVRMRPGMYIGSTSKEGLHHLVWEIVDNSIDEA 58

Query: 61  LAGFAGHIKVYIEPDNSITVVDDGRGIPVDIQEKTGRPAVETVFTVLHAGGKFGGGGYKV 120
           LAGFA HI+V+IEPD+SITVVDDGRGIPVDIQEKTGRPAVETVFTVLHAGGKFGGGGYKV
Sbjct: 59  LAGFASHIQVFIEPDDSITVVDDGRGIPVDIQEKTGRPAVETVFTVLHAGGKFGGGGYKV 118

Query: 121 SGGLHGVGSSVVNALSTQLDVKVYKNGKVHYQEYQRGVVVNDLEIIGDTDLSGTTVHFTP 180
           SGGLHGVGSSVVNALSTQLDV V+KNGK+HYQEY+RG VV DLE++GDTD +GTTVHFTP
Sbjct: 119 SGGLHGVGSSVVNALSTQLDVHVHKNGKIHYQEYRRGHVVADLEVVGDTDRTGTTVHFTP 178

Query: 181 DPEIFTETTVFDFDKLAKRIQELAFLNRGLRISISDKREGQEVEKEYHYEGGIGSYVEFI 240
           DPEIFTETT+FDFDKL KRIQELAFLNRGL+ISI+DKR+G E  K YHYEGGI SYVE+I
Sbjct: 179 DPEIFTETTIFDFDKLNKRIQELAFLNRGLQISITDKRQGLEQTKHYHYEGGIASYVEYI 238

Query: 241 NENKEVIFENPIYTDGELDGISVEVAMQYTTGYQETVMSFANNIHTHEGGTHEQGFRTAL 300
           NENK+VIF+ PIYTDGE+D I+VEVAMQYTTGY E VMSFANNIHTHEGGTHEQGFRTAL
Sbjct: 239 NENKDVIFDTPIYTDGEMDDITVEVAMQYTTGYHENVMSFANNIHTHEGGTHEQGFRTAL 298

Query: 301 TRVINDYAKKNKILKENEDNLTGEDVREGLTAVISVKHPNPQFEGQTKTKLGNSEVVKIT 360
           TRVINDYA+KNK+LK+NEDNLTGEDVREGLTAVISVKHPNPQFEGQTKTKLGNSEVVKIT
Sbjct: 299 TRVINDYARKNKLLKDNEDNLTGEDVREGLTAVISVKHPNPQFEGQTKTKLGNSEVVKIT 358

Query: 361 NRLFSEAFNRFLLENPQVAKKIVEKGILASKARIAAKRAREVTRKKSGLEISNLPGKLAD 420
           NRLFSEAF+ FL+ENPQ+AK+IVEKGILA+KAR+AAKRAREVTRKKSGLEISNLPGKLAD
Sbjct: 359 NRLFSEAFSDFLMENPQIAKRIVEKGILAAKARVAAKRAREVTRKKSGLEISNLPGKLAD 418

Query: 421 CSSNNAEMNELFIVEGDSAGGSAKSGRNREFQAILPIRGKILNVEKATMDKILANEEIRS 480
           CSSNN    ELFIVEGDSAGGSAKSGRNREFQAILPIRGKILNVEKA+MDKILANEEIRS
Sbjct: 419 CSSNNPAETELFIVEGDSAGGSAKSGRNREFQAILPIRGKILNVEKASMDKILANEEIRS 478
```

```
Query: 481 LFTAMGTGFGADFDVSKVRYQKLVIMTDADVDGAHIRTLLLTLIYRFMRPVLEAGYVYIA 540
           LFTAMGTGFGA+FDVSK RYQKLV+MTDADVDGAHIRTLLLTLIYR+M+P+LEAGYVYIA
Sbjct: 479 LFTAMGTGFGAEFDVSKARYQKLVLMTDADVDGAHIRTLLLTLIYRYMKPILEAGYVYIA 538

Query: 541 QPPIYGVKVGSEIKAYIQPGVNQEEELRQALDTYSSGRSKPTVQRYKGLGEMDDHQLWET 600
           QPPIYGVKVGSEIK YIQPG +QE +L++AL  YS GR+KPT+QRYKGLGEMDDHQLWET
Sbjct: 539 QPPIYGVKVGSEIKEYIQPGADQEIKLQEALARYSEGRTKPTIQRYKGLGEMDDHQLWET 598

Query: 601 TMDPENRLMARVSVDDAAEADKIFDMLMGDRVEPRREFIEANAVYSNLDI 650
           TMDPE+RLMARVSVDDAAEADKIFDMLMGDRVEPRREFIE NAVYS LD+
Sbjct: 599 TMDPEHRLMARVSVDDAAEADKIFDMLMGDRVEPRREFIEENAVYSTLDV 648
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1317> which encodes the amino acid sequence <SEQ ID 1318>. Analysis of this protein sequence reveals the following:

```
Possible site: 41
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1698(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 584/650 (89%), Positives = 618/650 (94%)

Query: 1   MTEETKNMEQRAQEYDASQIQVLEGLEAVRMRPGMYIGSTSKEGLHHLVWEIVDNSIDEA 60
           M EE K+ E++ QEYDASQIQVLEGLEAVRMRPGMYIGST+KEGLHHLVWEIVDNSIDEA
Sbjct: 1   MIEENKHFEKKMQEYDASQIQVLEGLEAVRMRPGMYIGSTAKEGLHHLVWEIVDNSIDEA 60

Query: 61  LAGFAGHIKVYIEPDNSITVVDDGRGIPVDIQEKTGRPAVETVFTVLHAGGKFGGGGYKV 120
           LAGFA HIKV+IE DNSITVVDDGRGIPVDIQ KTGRPAVETVFTVLHAGGKFGGGGYKV
Sbjct: 61  LAGFASHIKVFIEADNSITVVDDGRGIPVDIQAKTGRPAVETVFTVLHAGGKFGGGGYKV 120

Query: 121 SGGLHGVGSSVVNALSTQLDVKVYKNGKVHYQEYQRGVVVNDLEIIGDTDLSGTTVHFTP 180
           SGGLHGVGSSVVNALSTQLDV+VYKNG++HYQE++RG VV DLE+IG TD++GTTVHFTP
Sbjct: 121 SGGLHGVGSSVVNALSTQLDVRVYKNGQIHYQEFKRGAVVADLEVIGTTDVTGTTVHFTP 180

Query: 181 DPEIFTETTVFDFDKLAKRIQELAFLNRGLRISISDKREGQEVEKEYHYEGGIGSYVEFI 240
           DPEIFTETT FD+  LAKRIQELAFLNRGL+ISI+DKR G E E+ + YEGGIGSYVEF+
Sbjct: 181 DPEIFTETTQFDYSVLAKRIQELAFLNRGLKISITDKRSGMEQEEHFLYEGGIGSYVEFL 240

Query: 241 NENKEVIFENPIYTDGELDGISVEVAMQYTTGYQETVMSFANNIHTHEGGTHEQGFRTAL 300
           N+ K+VIFE PIYTDGEL+GI+VEVAMQYTT YQETVMSFANNIHTHEGGTHEQGFR AL
Sbjct: 241 NDKKDVIFETPIYTDGELEGIAVEVAMQYTTSYQETVMSFANNIHTHEGGTHEQGFRAAL 300

Query: 301 TRVINDYAKKNKILKENEDNLTGEDVREGLTAVISVKHPNPQFEGQTKTKLGNSEVVKIT 360
           TRVINDYAKKNKILKENEDNLTGEDVREGLTAVISVKHPNPQFEGQTKTKLGNSEVVKIT
Sbjct: 301 TRVINDYAKKNKILKENEDNLTGEDVREGLTAVISVKHPNPQFEGQTKTKLGNSEVVKIT 360

Query: 361 NRLFSEAFNRFLLENPQVAKKIVEKGILASKARIAAKRAREVTRKKSGLEISNLPGKLAD 420
           NRLFSEAF RFLLENPQVA+KIVEKGILASKARIAAKRAREVTRKKSGLEISNLPGKLAD
Sbjct: 361 NRLFSEAFQRFLLENPQVARKIVEKGILASKARIAAKRAREVTRKKSGLEISNLPGKLAD 420

Query: 421 CSSNNAEMNELFIVEGDSAGGSAKSGRNREFQAILPIRGKILNVEKATMDKILANEEIRS 480
           CSSN+A  NELFIVEGDSAGGSAKSGRNREFQAILPIRGKILNVEKATMDKILANEEIRS
Sbjct: 421 CSSNDANQNELFIVEGDSAGGSAKSGRNREFQAILPIRGKILNVEKATMDKILANEEIRS 480

Query: 481 LFTAMGTGFGADFDVSKVRYQKLVIMTDADVDGAHIRTLLLTLIYRFMRPVLEAGYVYIA 540
           LFTAMGTGFGADFDVSK RYQKLVIMTDADVDGAHIRTLLLTLIYRFMRPVLEAGYVYIA
Sbjct: 481 LFTAMGTGFGADFDVSKARYQKLVIMTDADVDGAHIRTLLLTLIYRFMRPVLEAGYVYIA 540

Query: 541 QPPIYGVKVGSEIKAYIQPGVNQEEELRQALDTYSSGRSKPTVQRYKGLGEMDDHQLWET 600
           QPPIYGVKVGSEIK YIQPG++QE++L+ AL+ YS GRSKPTVQRYKGLGEMDDHQLWET
Sbjct: 541 QPPIYGVKVGSEIKEYIQPGIDQEDQLKTALEKYSIGRSKPTVQRYKGLGEMDDHQLWET 600
```

```
Query: 601 TMDPENRLMARVSVDDAAEADKIFDMLMGDRVEPRREFIEANAVYSNLDI 650
           TMDPENRLMARV+VDDAAEADK+FDMLMGDRVEPRR+FIE NAVYS LDI
Sbjct: 601 TMDPENRLMARVTVDDAAEADKVFDMLMGDRVEPRRDFIEENAVYSTLDI 650
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 407

A DNA sequence (GBSx0442) was identified in *S.agalactiae* <SEQ ID 1319> which encodes the amino acid sequence <SEQ ID 1320>. Analysis of this protein sequence reveals the following:

```
Possible site: 26
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3186(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAA91552 GB:Z67740 unidentified [Streptococcus pneumoniae]
 Identities = 82/142 (57%), Positives = 105/142 (73%)

Query: 45  LKESTADAIAYFIPEEADFLKEYKANEAKVLETPILFQGAKELLAKIQRQGSRNFLVSHR 104
            LK ST  AI  F P   +FL++YK NEA+ LE PILF+G  +LL  I  QG R+FLVSHR
Sbjct: 2   LKVSTPFAIETFAPNLENFLEKYKENEARELEHPILFEGVSDLLEDILNQGGRHFLVSHR 61

Query: 105 DNQVIVILEKTEIIDYFTEVVTADNGFSRKPSPESMLYLKEKYQIDNCLVIGDRDIDKQA 164
            ++QV+ ILEKT I  YFTEVVT+ +GF RKP+PESMLYL+EKYQI + LVIGDR ID +A
Sbjct: 62  NDQVLEILEKTSIAAYFTEVVTSSSGFKRKPNPESMLYLREKYQISSGLVIGDRPIDIEA 121

Query: 165 GESAGFDTLLVDGSKSLMEIIE 186
            G++AG DT L     +L ++++
Sbjct: 122 GQAAGLDTHLFTSIVNLRQVLD 143
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1321> which encodes the amino acid sequence <SEQ ID 1322>. Analysis of this protein sequence reveals the following:

```
Possible site: 59
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2472(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 122/185 (65%), Positives = 145/185 (77%)

Query: 1   MNYHDYIWDLGGTLLDNYESSTRAFVETLKEFGYQADHDSVYQKLKESTADAIAYFIPEE 60
            MNY DYIWDLGGTLLDNYE ST+AFV+TL  F    DHD+VYQKLKESTA A+A F P E
Sbjct: 4   MNYQDYIWDLGGTLLDNYELSTQAFVQTLAFFSLPGDHDAVYQKLKESTAIAVAMFAPNE 63

Query: 61  ADFLKEYKANEAKVLETPILFQGAKELLAKIQRQGSRNFLVSHRDNQVIVILEKTEIIDY 120
             +FL  Y+  EA  L  PI   GAKE+L KI   GSRNFL+SHRD QV  +LE+  ++ Y
Sbjct: 64  PEFLHVYRLREADKLAQPIWCLGAKEILGKIATSGSRNFLISHRDCQVNQLLEQAGLLIY 123

Query: 121 FTEVVTADNGFSRKPSPESMLYLKEKYQIDNCLVIGDRDIDKQAGESAGFDTLLVDGSKS 180
            FTEVVTA NGF+RKP+PES+ YLKEKY I++ LVIGDR IDKQAG++AGF+TLLVDG K+
Sbjct: 124 FTEVVTASNGFARKPNPESLFYLKEKYDINSGLVIGDRLIDKQAGQAAGFNTLLVDGRKN 183
```

```
Query: 181 LMEII 185
           L+EI+
Sbjct: 184 LLEIV 188
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 408

A DNA sequence (GBSx0443) was identified in *S.agalactiae* <SEQ ID 1323> which encodes the amino acid sequence <SEQ ID 1324>. This protein is predicted to be stage V sporulation protein E (rodA). Analysis of this protein sequence reveals the following:

```
Possible site: 42
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood =-11.15   Transmembrane  206 - 222 ( 177 - 226)
   INTEGRAL    Likelihood =-10.14   Transmembrane   58 -  74 (  50 -  82)
   INTEGRAL    Likelihood = -9.34   Transmembrane  182 - 198 ( 177 - 205)
   INTEGRAL    Likelihood = -8.55   Transmembrane  158 - 174 ( 156 - 177)
   INTEGRAL    Likelihood = -8.12   Transmembrane  300 - 316 ( 299 - 324)
   INTEGRAL    Likelihood = -2.66   Transmembrane   86 - 102 (  83 - 102)
   INTEGRAL    Likelihood = -2.34   Transmembrane  338 - 354 ( 338 - 357)

----- Final Results -----
               bacterial membrane --- Certainty=0.5458(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9669> which encodes amino acid sequence <SEQ ID 9670> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB15838 GB:Z99123 alternate gene name: ipa-42d~similar to
           cell-division protein [Bacillus subtilis]
 Identities = 142/392 (36%), Positives = 237/392 (60%), Gaps = 23/392 (5%)

Query: 10  QKSNYFKGQIDYAVVIPVFFLLMIGLASIYVA-TMNDYPSNIYIAMFQQVSWIIMGCIIA 68
            Q+S +++G  D   +  VFF+  I + SIY A     Y +  +I   QQ+ + ++G +
Sbjct: 7   QQSPFYQG--DLIFIFGVFFI--ISVVSIYAAGQFGQYGNTDWI---QQIVFYLLGAVAI 59

Query: 69  FVVMLFSTEFLWKATPYLYALGLTLMVLPLIFYSPQLFAAT--GAKNWVTIGSVTLFQPS 126
             V++ F  E L K + Y++ +G+  +++  I  SP+  A    GAK+W  IG +T+ QPS
Sbjct: 60  TVLLYFDLEQLEKLSLYIFIIGILSLIILKI--SPESIAPVIKGAKSWFRIGRITI-QPS 116

Query: 127 EFMKISYILMLSRITVSFHQKNRKTFQDDWKLL-GLFGLVTLPVMILLMLQKDLGTALVF 185
            EFMK+  I+ML+ +    + K  +T +DD  LL  + G+  +PV ++LM  +D GTA +
Sbjct: 117 EFMKVGLIMMLASVIGKANPKGVRTLRDDIHLLLKIAGVAVIPVGLILM--QDAGTAGIC 174

Query: 186 LAILSGLILLSGISWWIILPILSTIVLFIASFLMIFISPNGKEWFYNLGMDTYQINRLSA 245
            + I+  ++ +SGI+W +I  I  + +L I+  L++ I  N  +   ++G+  YQI R+++
Sbjct: 175 MFIVLVMVFMSGINWKLIAIIAGSGILLISLILLVMI--NFPDVAKSVGIQDYQIKRVTS 232

Query: 246 WIDPFSFAD---SIAYQQTQGMVSIGSGGVTGKGFNILELSVPVRESDMIFTVIAENFGF 302
            W+   +      + ++Q  Q +++IGSGG+ G G + L++ VP   +D IF++I E+FGF
Sbjct: 233 WVSASNETQEDSNDSWQVDQAIMAIGSGGILGNGISNLKVYVPESTTDFIFSIIGESFGF 292

Query: 303 IGSAIVLGLYLIIIYRMLRIT--IESNNQFYTFISTGFIMMIVFHVFENIGAAVGILPLT 360
            IG AIV+ ++  +IYR++ +   I   N+F +F   G+  +IV H F+NIG  +GI+P+T
Sbjct: 293 IGCAIVVIMFFFLIYRLVVLIDKIHPFNRFASFFCVGYTALIVIHTFQNIGMNIGIMPVT 352

Query: 361 GIPLPFISQGGSSLLSNLIGIGLVLSMSYQNT 392
            GIPL F+S GGSS LS LIG G+V + S Q T
Sbjct: 353 GIPLLFVSYGGSSTLSTLIGFGIVYNASVQLT 384
```

There is also homology to SEQ ID 1028.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 409

A DNA sequence (GBSx0444) was identified in *S.agalactiae* <SEQ ID 1325> which encodes the amino acid sequence <SEQ ID 1326>. Analysis of this protein sequence reveals the following:

```
Possible site: 25
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3195(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1327> which encodes the amino acid sequence <SEQ ID 1328>. Analysis of this protein sequence reveals the following:

```
Possible site: 22
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2735(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 38/55 (69%), Positives = 48/55 (87%)

Query: 8  DEFKEAIDKGYISGNTVAIVRKNGKIFDYVLLHEEVREEEVVTVERVLDVLRKLS 62
           DEFK+AID GYI+G+TVAIVRK+G+IFDYVL HE+V+  EVVT E+V +VL +LS
Sbjct: 5  DEFKQAIDNGYIAGDTVAIVRKDGQIFDYVLPHEKVKNGEVVTKEKVEEVLVELS 59
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 410

A DNA sequence (GBSx0445) was identified in *S.agalactiae* <SEQ ID 1329> which encodes the amino acid sequence <SEQ ID 1330>. Analysis of this protein sequence reveals the following:

```
Possible site: 41
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4241(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1331> which encodes the amino acid sequence <SEQ ID 1332>. Analysis of this protein sequence reveals the following:

```
Possible site: 41
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4551(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 57/66 (86%), Positives = 63/66 (95%)

Query: 1  MSQEKLKSKLDQAKGGAKEGFGKITGDKELEAKGFIEKTIAKGKELADDAKDAVEGAVDA 60
           MS+EKLKSK++QA GG KEG GK+TGDKELEAKGF+EKTIAKGKELADDAK+AVEGAVDA
Sbjct: 1  MSEEKLKSKIEQASGGLKEGAGKLTGDKELEAKGFVEKTIAKGKELADDAKEAVEGAVDA 60

Query: 61 VKEKLK 66
           VKEKLK
Sbjct: 61 VKEKLK 66
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 411

A DNA sequence (GBSx0447) was identified in *S.agalactiae* <SEQ ID 1333> which encodes the amino acid sequence <SEQ ID 1334>. This protein is predicted to be TnpA (orfB). Analysis of this protein sequence reveals the following:

```
Possible site: 16
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3961(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9667> which encodes amino acid sequence <SEQ ID 9668> was also identified.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1335> which encodes the amino acid sequence <SEQ ID 1336>. Analysis of this protein sequence reveals the following:

```
Possible site: 55
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3365(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 152/160 (95%), Positives = 154/160 (96%)

Query: 1    MKNMALPKMATVKTKTALKKTQKTYPQNLLNQKFNPDKPNQVWSTDFTYISIGYKKYVYL 60
            MKNMALPKMATVK KTALK+TQKTYPQNLLNQKFNPDKPNQVWSTDFTYISIGYKKYVYL
Sbjct: 194 MKNMALPKMATVKPKTALKRTQKTYPQNLLNQKFNPDKPNQVWSTDFTYISIGYKKYVYL 253

Query: 61   CAIIDLYSRKYIAWKLSHRMDAKLACDTLELALNKRKIEGTLLFHSDQGSQFKAREFRKI 120
            CAI+DLYSRK IAWKLSHRMDAKLACDTLELALNKRKIEGTLLFHSDQGSQFKARE RKI
Sbjct: 254 CAILDLYSRKCIAWKLSHRMDAKLACDTLELALNKRKIEGTLLFHSDQGSQFKARELRKI 313
```

```
Query: 121 IDDNNIMHSFSKPRYPYDNAVTEAFFKYLKHRQINQKNYQ 160
           IDDN IMHSFSKP YPYDNAVTEAFFKYLKHRQINQK YQ
Sbjct: 314 IDDNTIMHSFSKPGYPYDNAVTEAFFKYLKHRQINQKKYQ 353
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 412

A DNA sequence (GBSx0448) was identified in *S.agalactiae* <SEQ ID 1337> which encodes the amino acid sequence <SEQ ID 1338>. Analysis of this protein sequence reveals the following:

```
Possible site: 27
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1090(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 413

A DNA sequence (GBSx0449) was identified in *S.agalactiae* <SEQ ID 1339> which encodes the amino acid sequence <SEQ ID 1340>. This protein is predicted to be histidine kinase (resE). Analysis of this protein sequence reveals the following:

```
Possible site: 40
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-11.57    Transmembrane    17 -  33 (   6 -  38)
   INTEGRAL    Likelihood = -4.67    Transmembrane   147 - 163 ( 142 - 166)

----- Final Results -----
               bacterial membrane --- Certainty=0.5628(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAD25109 GB:AF140356 VncS [Streptococcus pneumoniae]
 Identities = 178/435 (40%), Positives = 281/435 (63%), Gaps = 1/435 (0%)

Query: 1   MKKLKIFPKMFIQIFSILGILIILVHSLFFFIFPKTYLETRKVKIHIMADEISKNMNGKE 60
           MK+  +F K+FI  FSI  +L+I +H   +F+FP TYL  R+  I   A  I++++ GK+
Sbjct: 1   MKRTGLFAKIFIYTFSIFSVLVICLHLAIYFLFPSTYLSHRQETIGQKATAIAQSLEGKD 60

Query: 61  LKYLDQTLELYSKSSDIKVFIKKNNNKNELQINDNINVNVKSDSNSLIIEEREIKLHDGK 120
            + ++Q L+LYS++SDIK  +K    +++L++ D++ ++     + SL IEERE+K  DG
Sbjct: 61  RQSIEQVLDLYSQTSDIKGTVKGEMTEDKLEVKDSLPLDTDRQTTSLFIEEREVKTQDGG 120

Query: 121 KIHLQFVSTADMQKDAKDLSLKFLPYSLSISFLFSIVISLIYAKSIKNNIQEITMVTDKM 180
            + LQF+++ D+QK+A+ +SL+FLPY+L  SFL S++++ IYA++I   I EI  VT +M
Sbjct: 121 TMILQFLASMDLQKEAEQISLQFLPYTLLASFLISLLVAYIYARTIVAPILEIKRVTRRM 180

Query: 181 IKLDKETRLKISSNDEIGQLKQQINDLYCALLNTINDLEFKNKEILKLEKLKYDFFKGAS 240
           + LD + RL++ S DEIG LK+QIN LY  LL  I DL  KN+ IL+LEK+K +F +GAS
Sbjct: 181 MDLDSQVRLRVDSKDEIGNLKEQINSLYQHLLTVIADLHEKNEAILQLEKMKVEFLRGAS 240
```

```
Query: 241 HELKTPLSSLKILLENMKYNIGKYKDRDFYISECINIVDNLTKNVSQILSFYSIKDLNND 300
           HELKTPL+SLKIL+ENM+ NIG+YKDRD Y+   + IVD L  +V QILS  S+++L +D
Sbjct: 241 HELKTPLASLKILIENMRENIGRYKDRDQYLGVALGIVDELNHHVLQILSLSSVQELRDD 300

Query: 301 EEYLNVGDTLDEVLEKYSILVNQKKININKELLDYNIYIGKTALNIVFSNLISNAVKYTN 360
            E +++      +++ Y++L  ++++ I+  L     Y+  + + ++ SNLISNA+K++
Sbjct: 301 RETIDLLQMTQNLVKDYALLAKERELQIDNSLTHQQAYLNPSVMKLILSNLISNAIKHSV 360

Query: 361 RNGIININKIANDWLLIENSYDKNKISKINKILDASFDLKLDNSNGLGLNIVKNILNKYNI 420
             G++ I      L IENS    +  K+ +    +   K+  S G+GL +VK++L    +
Sbjct: 361 PGGLVRIGEREGELFIENSCSSEEQEKLAQSFSDNASRKVKGS-GMGLFVVKSLLEHEKL 419

Query: 421 KYEILHGENYFIFKI 435
            Y     EN   F I
Sbjct: 420 AYRFEMEENSLTFFI 434
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1341> which encodes the amino acid sequence <SEQ ID 1342>. Analysis of this protein sequence reveals the following:

```
        Possible site: 37

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-11.83    Transmembrane    14 -  30 (   6 -  35)
   INTEGRAL    Likelihood = -2.44    Transmembrane   157 - 173 ( 156 - 174)

----- Final Results -----
               bacterial membrane --- Certainty=0.5734(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAD25109 GB:AF140356 VncS [Streptococcus pneumoniae]
 Identities = 123/455 (27%), Positives = 223/455 (48%), Gaps = 23/455 (5%)

Query: 3   LIKKTFLVINGLIIVVVTSILLVLYFAMPIYYTKVKDKEVKCEFDQTSKQIKGKTVTEIR 62
           L  K F+    +  V+V  + L +YF  P  Y   + + +  +    ++ ++GK    I
Sbjct: 6   LFAKIFIYTFSIFSVLVICLHLAIYFLFPSTYLSHRQETIGQKATAIAQSLEGKDRQSIE 65

Query: 63  DILTKKINKDNIWYSLVDSDNQLLYPSLQLLDGVSESKDSQNVNIVTTFDNSYSNVKVMS 122
            +L       +I  ++     ++    L++ D +    D Q  ++            +
Sbjct: 66  QVLDLYSQTSDIKGTV---KGEMTEDKLEVKDSLPLDTDRQTTSLF-----------IEE 111

Query: 123 QKVTLRDGKKMTLLGQSSLQPVTDASKVLLDLYPSLLIFSVTVGSIVAYLYSRTSSRRIL 182
           ++V  +DG  M L   +S+    +A ++ L   P  L+ S  +  +VAY+Y+RT    IL
Sbjct: 112 REVKTQDGGTMILQFLASMDLQKEAEQISLQFLPYTLLASFLISLLVAYIYARTIVAPIL 171

Query: 183 SMSQTAKKMVNLEPNLTCTIHGKDEIAMLASDINRLYASLSTSIKSLQKEYEKASDSERE 242
            + +  ++M++L+  +   +  KDEI  L   IN LY  L T I  L ++ E     E+
Sbjct: 172 EIKRVTRRMMDLDSQVRLRVDSKDEIGNLKEQINSLYQHLLTVIADLHEKNEAILQLEKM 231

Query: 243 KSEFLRMTSHELKTPITSVIGMIDGMLYNVGDFADRDKYLRKCRDVLEGQAQLVQSILSL 302
           K EFLR  SHELKTP+ S+  +I+ M  N+G + DRD+YL     +++     V  ILSL
Sbjct: 232 KVEFLRGASHELKTPLASLKILIENMRENIGRYKDRDQYLGVALGIVDELNHHVLQILSL 291

Query: 303 SKIETLASQNQELFSLKSSLEEEMEVFLVLSELKHLKVTINLEEQFVKANKVYLLKAIKN 362
           S ++ L   ++E   L    +  ++ + +L++ + L++  +L  Q    N   +   + N
Sbjct: 292 SSVQEL-RDDRETIDLLQMTQNLVKDYALLAKERELQIDNSLTHQQAYLNPSVMKLILSN 350

Query: 363 IIDNAFHYTKSGGQVMIQLKDNQLVIKNEAETLLTQQQMKQLFQPFYRPDYSRNRKDGGT 422
           +I NA  ++  GG V I  ++ +L I+N      + ++ ++L Q F     + +RK  G+
Sbjct: 351 LISNAIKHSVPGGLVRIGEREGELFIENSC----SSEEQEKLAQSF---SDNASRKVKGS 403

Query: 423 GLGLFITHQILDQHHLAYRFVVLDQRWMVFTIDFP 457
           G+GLF+   +L+   LAYRF  +++  + F IDFP
Sbjct: 404 GMGLFVVKSLLEHEKLAYRF-EMEENSLTFFIDFP 437
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 108/454 (23%), Positives = 220/454 (47%), Gaps = 22/454 (4%)

Query: 4   LKIFPKMFIQIFSILGILIILVHSLFFFIFPKTYLETRKVKIHIMADEISKNMNGKELKY 63
           +++  K F+ I  ++ +++  +  + +F  P  Y + +  ++    D+ SK + GK +
Sbjct: 1   VRLIKKTFLVINGLIIVVVTSILLVLYFAMPIYYTKVKDKEVKCEFDQTSKQIKGKTVTE 60

Query: 64  LDQTLELYSKSSDIKVFIKKNNNK------------NELQINDNINVNVKSDSN--SLII 109
           +   L       +I   +  ++N+            +E + + N+N+    D++  ++ +
Sbjct: 61  IRDILTKKINKDNIWYSLVDSDNQLLYPSLQLLDGVSESKDSQNVNIVTTFDNSYSNVKV 120

Query: 110 EEREIKLHDGKKIHLQFVSTADMQKDAKDLSLKFLPYSLSISFLFSIVISLIYAKSIKNN 169
             +++ L DGKK+ L   S+     DA  + L   P  L  S     +++ +Y+++
Sbjct: 121 MSQKVTLRDGKKMTLLGQSSLQPVTDASKVLLDLYPSLLIFSVTVGSIVAYLYSRTSSRR 180

Query: 170 IQEITMVTDKMIKLDKETRLKISSNDEIGQLKQQINDLYCALLNTINDLEFKNKEILKLE 229
           I  ++    KM+ L+      I   DEI  L   IN LY +L  +I  L+ + ++    E
Sbjct: 181 ILSMSQTAKKMVNLEPNLTCTIHGKDEIAMLASDINRLYASLSTSIKSLQKEYEKASDSE 240

Query: 230 KLKYDFFKGASHELKTPLSSLKILLENMKYNIGKYKDRDFYISECINIVDNLTKNVSQIL 289
           + K +F +  SHELKTP++S+  +++ M YN+G + DRD Y+ +C ++++   + V  IL
Sbjct: 241 REKSEFLRMTSHELKTPITSVIGMIDGMLYNVGDFADRDKYLRKCRDVLEGQAQLVQSIL 300

Query: 290 SFYSIKDL-NNDEEYLNVGDTLDEVLEKYSILVNQKKININKELLDYNIYIGKTALNIVF 348
           S   I+ L + ++E  ++  +L+E +E + +L   K + +   L +  +   K  L
Sbjct: 301 SLSKIETLASQNQELFSLKSSLEEEMEVFLVLSELKHLKVTINLEEQFVKANKVYLLKAI 360

Query: 349 SNLISNAVKYTNRNGIINIKIANDWLLIENSYDKNKISKINKILDASF------DLKLDN 402
            N+I NA  YT   G + I++ ++ L+I+N  +     +  K L   F        + D
Sbjct: 361 KNIIDNAFHYTKSGGQVMIQLKDNQLVIKNEAETLLTQQQMKQLFQPFYRPDYSRNRKDG 420

Query: 403 SNGLGLNIVKNILNKYNIKYE-ILHGENYFIFKI 435
             GLGL I   IL+++++ Y  ++  + + +F I
Sbjct: 421 GTGLGLFITHQILDQHHLAYRFVVLDQRWMVFTI 454
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 414

A DNA sequence (GBSx0450) was identified in *S.agalactiae* <SEQ ID 1343> which encodes the amino acid sequence <SEQ ID 1344>. This protein is predicted to be response regulator (regX3). Analysis of this protein sequence reveals the following:

```
Possible site: 34
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.80   Transmembrane   50 -   66 (  50 -   66)

----- Final Results -----
              bacterial membrane --- Certainty=0.1319(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9665> which encodes amino acid sequence <SEQ ID 9666> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAD25108 GB:AF140356 VncR [Streptococcus pneumoniae]
 Identities = 131/218 (60%), Positives = 176/218 (80%), Gaps = 1/218 (0%)

Query: 5   MKILTVEDDKLIREGISEYLSEFGYTVIQAKDGREALSKFNS-DINLVILDIQIPFINGL 63
```

```
           MKIL VED+++IREG+S+YL++ GY  I+A DG+EAL +F+S ++ LV+LDIQ+P +NGL
Sbjct: 1   MKILIVEDEEMIREGVSDYLTDCGYETIEAADGQEALEQFSSYEVALVLLDIQMPKLNGL 60

Query: 64  EVLKEIRKKSNLPILILTAFSDEEYKIDAFTNLVDGYVEKPFSLPVLKARIDSLIKKNFG 123
           EVL EIRK S +P+L+LTAF DEEYK+ AF +L DGY+EKPFSL +LK R+D++ K+ +
Sbjct: 61  EVLAEIRKTSQVPVLMLTAFQDEEYKMSAFASLADGYLEKPFSLSLLKVRVDAIFKRYYD 120

Query: 124 HLEKFEYKNLSVNFNSYTAKINDEKIDVNAKELEILKCLLDNDGQVLTRMQIIDYVWKDS 183
               F YK+  V+F SY+A +  +++ +NAKELEIL  L+ N+G+ LTR QIID VWK +
Sbjct: 121 TGRIFSYKDTKVDFESYSASLAGQEVPINAKELEILDYLVKNEGRALTRSQIIDAVWKAT 180

Query: 184 EEIPYDRVVDVYIKELRKKLQLDCITTIRNVGYKLERK 221
           +E+P+DRV+DVYIKELRKKL LDCI T+RNVGYKLERK
Sbjct: 181 DEVPFDRVIDVYIKELRKKLDLDCILTVRNVGYKLERK 218
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1345> which encodes the amino acid sequence <SEQ ID 1346>. Analysis of this protein sequence reveals the following:

```
        Possible site: 60
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -2.60    Transmembrane   48 -   64 (   48 -   64)

----- Final Results -----
               bacterial membrane --- Certainty=0.2041(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAF72358 GB:AF192329 VanRB [Enterococcus faecalis]
 Identities = 88/215 (40%), Positives = 128/215 (58%), Gaps = 2/215 (0%)

Query: 3   KILVVEDDDTISQVICEFLKANNYDPDCVFDGQAALDKWQTTSYDLIILDIMLPSLSGLE 62
           +IL+VEDDD I   +  FL    Y  D   DG  A  K+   +Y L+ILDIMLP ++G E
Sbjct: 4   RILLVEDDDHICNTVRGFLAEAGYQVDACTDGNEAYTKFYENTYQLVILDIMLPGMNGHE 63

Query: 63  VLKTIRKTSDVPIIMLTALDDEYTQLVSFNHLISDYVTKPFSPLILIKRIENVLRVSTPD 122
           +L+  R  +D PI+M+TAL D+  Q+ +F+    DYVTKPF   IL+KR+E +LR S
Sbjct: 64  LLREFRAKNDTPILMMTALSDDENQIRAFDAEADDYVTKPFKMQILLKRVEALLRRSGAL 123

Query: 123 EKR-QIGDLLVDETEHSVYWQGTLVKLTKKEYDIIDYLAKRHQKIVTRDQLMDDIWGYS- 180
            K  ++G L +   + +V   GT + LT+KE++I+  L +   + +T + ++  IWGY
Sbjct: 124 AKEIRVGRLTLLPEDFTVLCDGTELPLTRKEFEILLLLVQNKGRTLTHEIILSRIWGYDF 183

Query: 181 ELDTRVLDNHIKNLRKKMTGIPLKTITGMGYLLGE 215
           E D   +  HIKNLR K+    +KTI G+GY L E
Sbjct: 184 EGDGSTVHTHIKNLRAKLPENIIKTIRGVGYRLEE 218
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 80/214 (37%), Positives = 126/214 (58%), Gaps = 4/214 (1%)

Query: 6   KILTVEDDKLIREGISEYLSEFGYTVIQAKDGREALSKFNS-DINLVILDIQIPFINGLE 64
           KIL VEDD  I + I E+L    Y      DG+ AL K+ +   +L+ILDI +P ++GLE
Sbjct: 3   KILVVEDDDTISQVICEFLKANNYDPDCVFDGQAALDKWQTTSYDLIILDIMLPSLSGLE 62

Query: 65  VLKEIRKKSNLPILILTAFSDEEYKIDAFTNLVDGYVEKPFSLPVLKARIDSLIKKNFGH 124
           VLK IRK S++PI++LTA  DE  ++ +F +L+  YV KPFS  +L  RI+++++ +
Sbjct: 63  VLKTIRKTSDVPIIMLTALDDEYTQLVSFNHLISDYVTKPFSPLILIKRIENVLRVSTPD 122

Query: 125 LEKFEYKNLSVNFNSYTAKINDEKIDVNAKELEILKCLLDNDGQVLTRMQIIDYVWKDSE 184
            EK +  +L V+   ++       + +  KE +I+  L     +++TR Q++D +W  SE
Sbjct: 123 -EKRQIGDLLVDETEHSVYWQGTLVKLTKKEYDIIDYLAKRHQKIVTRDQLMDDIWGYSE 181

Query: 185 EIPYDRVVDVYIKELRKKLQLDCITTIRNVGYKL 218
                RV+D +IK LRKK+    + TI  +GY L
Sbjct: 182 --LDTRVLDNHIKNLRKKMTGIPLKTITGMGYLL 213
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 415

A DNA sequence (GBSx0451) was identified in *S.agalactiae* <SEQ ID 1347> which encodes the amino acid sequence <SEQ ID 1348>. This protein is predicted to be Vexp3. Analysis of this protein sequence reveals the following:

```
Possible site: 49
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-12.68   Transmembrane  423 - 439 ( 413 - 447)
   INTEGRAL    Likelihood =-10.67   Transmembrane   16 -  32 (  12 -  37)
   INTEGRAL    Likelihood = -9.77   Transmembrane  303 - 319 ( 301 - 326)
   INTEGRAL    Likelihood = -3.13   Transmembrane  343 - 359 ( 343 - 367)

----- Final Results -----
               bacterial membrane --- Certainty=0.6074(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAD47594 GB:AF140784 Vexp3 [Streptococcus pneumoniae]
 Identities = 280/458 (61%), Positives = 363/458 (79%), Gaps = 3/458 (0%)

Query: 1   MIKNAFAYVTRKSLKSLIIILVILSMATLSIISLSIKDATDRASKETFANITNSFSMEIN 60
           M+ NAFAYVTRK  KS++I L+IL MA+LS++ LSIK AT +AS+ETF NITNSFSM+IN
Sbjct: 1   MLHNAFAYVTRKFFKSIVIFLIILLMASLSLVGLSIKGATAKASQETFKNITNSFSMQIN 60

Query: 61  RQVNPGTPRGGGNVKGEDIKKISQTNSIDSYVKRINSVADLVDHDIIETQDTLANQSPER 120
           R+VN GTPRG GN+KGEDIKKI++  +I+SYVKRIN++ DL  +D+IET +T  N + +R
Sbjct: 61  RRVNQGTPRGAGNIKGEDIKKITENKAIESYVKRINAIGDLTGYDLIETPETKKNLTADR 120

Query: 121 AKNFKRTVMLTGVNDSAKETKFVSEAYKLVEGKHLENKDKNKILMHKDLAKKNNLKVGDK 180
           AK F  ++M+TGVNDS+KE KFVS +YKLVEG+HL N DK+KIL+HKDLA K+  KVGDK
Sbjct: 121 AKRFGSSLMITGVNDSSKEDKFVSGSYKLVEGEHLTNDDKDKILLHKDLAAKHGWKVGDK 180

Query: 181 IKIKSNLFDADNEKVANETVEVEIKGLFDGHNSGGVSAAQELYENTLITDVHSAAKVYGN 240
           +K+ SN++DADNEK A ETVEV IKGLFDGHN   V+ +QELYENT ITD+H+AAK+YG
Sbjct: 181 VKLDSNIYDADNEKGAKETVEVTIKGLFDGHNKSAVTYSQELYENTAITDIHTAAKLYGY 240

Query: 241 TEDTAVYQDATFFVKGDKNLDSVIKDL-GKLDINWREYNLIKSSSNYPALQQSISGIYSI 299
           TEDTA+Y DATFFV  DKNLD V+K+L G   INW+ Y L+KSSSNYPAL+QSISG+Y +
Sbjct: 241 TEDTAIYGDATFFVTADKNLDDVMKELNGISGINWKSYTLVKSSSNYPALEQSISGMYKM 300

Query: 300 SNKLFVGSLIFAGVVVSLLLFLWMNARKKEIAVLLSLGISKLEIFGQFIIEMVFISIPAL 359
           +N LF GSL F+ ++++LLL LW+NAR+KE+ +LLS+G+ +  I GQFI E + I+IPAL
Sbjct: 301 ANLLFWGSLSFSVLLLALLLSLWINARRKEVGILLSIGLKQASILGQFITESILIAIPAL 360

Query: 360 LGSYFLAQYTADKLGNNILNKVTGDIAKQIARQSASSQLGGGAEAEGFNKTLSGLDINV- 418
           + +YFLA YTA  +GN +L  VT  +AKQ ++ + +S LGGGAE +GF+KTLS LDI++
Sbjct: 361 VSAYFLANYTARAIGNTVLANVTSGVAKQASKAAQASNLGGGAEVDGFSKTLSSLDISIQ 420

Query: 419 LPKFIIYVVIFMSFVLLVSLILSSIYTLRKNPKELLID 456
              FII  V+ +  V+LV + L+S   LRK PKELL+D
Sbjct: 421 TSDFIIIFVLALVLVVLV-MALASSNLLRKQPKELLLD 457
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1349> which encodes the amino acid sequence <SEQ ID 1350>. Analysis of this protein sequence reveals the following:

```
     Possible site: 31
>>> Seems to have no N-terminal signal sequence
```

```
INTEGRAL    Likelihood =-12.90   Transmembrane   19 -  35 (  16 -  43)
INTEGRAL    Likelihood = -7.27   Transmembrane  371 - 387 ( 359 - 392)
INTEGRAL    Likelihood = -7.01   Transmembrane  335 - 351 ( 326 - 357)
INTEGRAL    Likelihood = -6.21   Transmembrane  282 - 298 ( 276 - 308)

----- Final Results -----
               bacterial membrane --- Certainty=0.6158(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAC24912 GB:AF012285 YknZ [Bacillus subtilis]
 Identities = 176/408 (43%), Positives = 250/408 (61%), Gaps = 16/408 (3%)

Query: 1   MENWKFALSSIWGHKMRSILTMLGIIIGVAAVVIIMGLGNAMKNSVTSTFSSKQKDIQLY 60
           +EN + ALSS+  HKMRSILTMLGIIIGV +V++++ +G   +  +  + S     ++LY
Sbjct: 4   LENIRMALSSVLAHKMRSILTMLGIIIGVGSVIVVVAVGQGGEQMLKQSISGPGNTVELY 63

Query: 61  FQEKGEE--EDLYAGLHTHENNHEVKPEWLEQIVKDIDGIDSYYFTNSATSTISYEKKKV 118
           +    EE   +  A   +    +++K       +K I+GI     + S +    Y +++
Sbjct: 64  YMPSDEELASNPNAAAESTFTENDIKG------LKGIEGIKQVVASTSESMKARYHEEET 117

Query: 119 DNASIIGVSKDYFNIKNYDIVAGRTLTDNDYSNFSRIILLDTVLADDLFGKGNYKSALNK 178
           D A++ G++  Y N+ +  I +GRT TDND+   +R+ ++   +A +LF K    S L +
Sbjct: 118 D-ATVNGINDGYMNVNSLKIESGRTFTDNDFLAGNRVGIISQKMAKELFDK---TSPLGE 173

Query: 179 VVSLSDKDYLVIGVYKTDQTPVSFDGLSGGAVMANTQVASEFGTKEIGSIYIHVNDIQNS 238
           VV ++ +   +IGV K     +SFD LS   V  N  + S FGT +  ++ + V     +
Sbjct: 174 VVWINGQPVEIIGVLKKVTGLLSFD-LSEMYVPFN-MMKSSFGTSDFSNVSLQVESADDI 231

Query: 239 MNLGNQAADMLTNISHIKDGQYAVPDNSKIVEEINSQFSIMTTVIGSIAAISLLVGGIGV 298
            + G +AA  L N +H  +  Y V +  +I   I    +IMTT+IGSIA ISLLVGGIGV
Sbjct: 232 KSAGKEAAQ-LVNDNHGTEDSYQVMNMEEIAAGIGKVTAIMTTIIGSIAGISLLVGGIGV 290

Query: 299 MNIMLVSVTERTREIGLRKALGATRLKILSQFLIESVVLTVLGGLIGLLLAQLSVGALGN 358
           MNIMLVSVTERTREIG+RK+LGATR +IL+QFLIESVVLT++GGL+G+ +      AL +
Sbjct: 291 MNIMLVSVTERTREIGIRKSLGATRGQILTQFLIESVVLTLIGGLVGIGIG-YGGAALVS 349

Query: 359 AMTLKGACISLDVALIAVLFSASIGVFFGMLPANKASKLDPIEALRYE 406
           A+    + IS  V    VLFS  IGV FGMLPANKA+KLDPIEALRYE
Sbjct: 350 AIAGWPSLISWQVVCGGVLFSMLIGVIFGMLPANKAAKLDPIEALRYE 397
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 56/247 (22%), Positives = 101/247 (40%), Gaps = 42/247 (17%)

Query: 147 YKLVEGKHLENKDKNKI--------LMHKDLAKKNNLK--------VGDKIKIKSNLFDA 190
           Y +V G+ L + D +          ++  DL  K N K        + DK  +   ++
Sbjct: 136 YDIVAGRTLTDNDYSNFSRIILLDTVLADDLFGKGNYKSALNKVVSLSDKDYLVIGVYKT 195

Query: 191 DNEKVANETVEVEIKGLFDGHNSGGVSAAQELYENTLITDVHSAAKVYGNTEDTAVYQDA 250
           D   V+           FDG + G V A      NT +      A  +G  E  ++Y
Sbjct: 196 DQTPVS-----------FDGLSGGAVMA------NTQV------ASEFGTKEIGSIYIHV 232

Query: 251 TFFVKGDKNLDSVIKDL--GKLDINWREYNLIKSSSNYPALQQSISGIYSISNKLFVGSL 308
              ++   NL +   D+      I   +Y +  +S     +    S + ++   +   SL
Sbjct: 233 ND-IQNSMNLGNQAADMLTNISHIKDGQYAVPDNSKIVEEINSQFSIMTTVIGSIAAISL 291

Query: 309 IFAGVVVSLLLFLWMNARKKEIAVLLSLGISKLEIFGQFIIEMVFISIPALLGSYFLAQY 368
           +  G+ V  ++ + +  R +EI +  +LG ++L+I  QF+IE V +++   L    LAQ
Sbjct: 292 LVGGIGVMNIMLVSVTERTREIGLRKALGATRLKILSQFLIESVVLTVLGGLIGLLLAQL 351

Query: 369 TADKLGN 375
           +   LGN
Sbjct: 352 SVGALGN 358
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 416

A DNA sequence (GBSx0452) was identified in *S.agalactiae* <SEQ ID 1351> which encodes the amino acid sequence <SEQ ID 1352>. This protein is predicted to be Vexp2 (b0879). Analysis of this protein sequence reveals the following:

```
Possible site: 16
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3194(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAD47593 GB:AF140784 Vexp2 [Streptococcus pneumoniae]
 Identities = 142/207 (68%), Positives = 169/207 (81%)

Query: 1   MDILEIKNVNYSYANSKEKVLSGVNQKFELGKFYAIVGKSGTGKSTLLSLLAGLDKVQTG 60
           M +L++++V Y Y N+ E VL  +N  FE GKFY+I+G+SG GKSTLLSLLAGLD    G
Sbjct: 1   MTLLQLQDVTYRYKNTAEAVLYQINYNFEPGKFYSIIGESGAGKSTLLSLLAGLDSPVEG 60

Query: 61  KILFKNEDIEKKGYSNHRKNNISLVFQNYNLIDYLSPIENIRLVNKSVDESILFELGLDK 120
            ILF+ EDI KKGYS HR ++ISLVFQNYNLIDYLSP+ENIRLVNK   ++ L ELGLD+
Sbjct: 61  SILFQGEDIRKKGYSYHRMHHISLVFQNYNLIDYLSPLENIRLVNKKASKNTLLELGLDE 120

Query: 121 KQIKRNVMKLSGGQQQRVAIARALVSDAPIILADEPTGNLDSVTAGEIINILKELAQDRN 180
            QIKRNV++LSGGQQQRVAIAR+LVS+AP+ILADEPTGNLD  TAG+I+ +LK LAQ
Sbjct: 121 SQIKRNVLQLSGGQQQRVAIARSLVSEAPVILADEPTGNLDPKTAGDIVELLKSLAQKTG 180

Query: 181 KCVIVVTHSKEVADSADIILELSGKKL 207
           KCVIVVTHSKEVA ++DI LEL  KKL
Sbjct: 181 KCVIVVTHSKEVAQASDITLELKDKKL 207
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1353> which encodes the amino acid sequence <SEQ ID 1354>. Analysis of this protein sequence reveals the following:

```
Possible site: 47
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2717(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 83/230 (36%), Positives = 135/230 (58%), Gaps = 13/230 (5%)

Query: 1   MDILEIKNVNYSYANSKEKVLSGVNQKFEL--GKFYAIVGKSGTGKSTLLSLLAGLDKVQ 58
           M  +E+K V+ SY   +  V +     FE+  G+   I+G SG GKST+L++L G+D V
Sbjct: 5   MAFIELKQVSKSYQIGETTVFANHEVSFEINKGELVVILGASGAGKSTVLNILGGMDTVD 64

Query: 59  TGKILFKNEDIE---KKGYSNHRKNNISLVFQNYNLIDYLSPIENIRLVNKSVDES---- 111
            G+++   +DI     K  + +R+N I  VFQ YNL+  L+  EN+ L  + V ++
Sbjct: 65  AGQVIIDGKDIAHYTSKALTQYRRNAIGFVFQFYNLVPNLTAKENVELAVEIVADALDPV 124

Query: 112 -ILFELGLDKKQIKRNVMKLSGGQQQRVAIARALVSDAPIILADEPTGNLDSVTAGEIIN 170
            IL E+GL  + +     +LSGG+QQRV+IARAL  +  ++L DEPTG LD  T  +I+
Sbjct: 125 TILKEVGLSHR-LDHFPAQLSGGEQQRVSIARALAKNPKLLLCDEPTGALDYQTGKQILT 183
```

```
Query: 171 ILKELAQDRNKCVIVVTHSKEVADSADIILELSGKKLKK--VNKMNLEVE 218
           +L+++AQ +   V++VTH+  +A  AD ++ +    ++ K  +NK    +E
Sbjct: 184 LLQDMAQTKGTTVVIVTHNAAIAPIADRVIFMHDAQVTKTVINKEPASIE 233
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 417

A DNA sequence (GBSx0453) was identified in *S.agalactiae* <SEQ ID 1355> which encodes the amino acid sequence <SEQ ID 1356>. Analysis of this protein sequence reveals the following:

```
Possible site: 25
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -3.35   Transmembrane   17 -  33 (  17 -  34)

----- Final Results -----
                  bacterial membrane --- Certainty=0.2338(Affirmative) < succ>
                   bacterial outside --- Certainty=0.0000(Not Clear) < succ>
                 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 418

A DNA sequence (GBSx0454) was identified in *S.agalactiae* <SEQ ID 1357> which encodes the amino acid sequence <SEQ ID 1358>. This protein is predicted to be Vexp1. Analysis of this protein sequence reveals the following:

```
Possible site: 56
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-11.52   Transmembrane  294 - 310 ( 285 - 312)
   INTEGRAL    Likelihood =-10.67   Transmembrane  396 - 412 ( 385 - 417)
   INTEGRAL    Likelihood = -8.76   Transmembrane   17 -  33 (  14 -  38)
   INTEGRAL    Likelihood = -4.14   Transmembrane  335 - 351 ( 333 - 357)

----- Final Results -----
                  bacterial membrane --- Certainty=0.5607(Affirmative) < succ>
                   bacterial outside --- Certainty=0.0000(Not Clear) < succ>
                 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAD47592 GB:AF140784 Vexp1 [Streptococcus pneumoniae]
 Identities = 165/425 (38%), Positives = 271/425 (62%), Gaps = 4/425 (0%)

Query: 2   IKNAIAYITRKKNRTLIIFAILTIVLSCLYSCLTIMKSSNEIEKALYESSNSSISITK-K 60
            I+ + AY++RK+ R+ I+F IL ++L+ + +CLT+MKS+  +E  LY+S N+S SI K +
Sbjct: 4   IQRSWAYVSRKRLRSFILFLILLVLLAGISACLTLMKSNKTVESNLYKSLNTSFSIKKIE 63

Query: 61  DGKYFNINQFKNIEKIKEVEEKIFQYDGLAKLKDLKVVSGEQSINREDLSDEFKNVVSLE 120
            +G+ F ++    ++ KIK +E    + + +AKLKD + V+GEQS+ R+DLS    N+VSL
Sbjct: 64  NGQTFKLSDLASVSKIKGLENVSPELETVAKLKDKEAVTGEQSVERDDLSAADNNLVSLT 123

Query: 121 ATSNTKRNLLFSSGVFSFKEGKNIEENDKNSILVHEEFAKQNKLKLGDEIDLELLDTEKS 180
            A  ++ +++ F+S  F+ KEG+++++ D    IL+HEE AK+N L L D+I L+    +E S
Sbjct: 124 ALEDSSKDVTFTSSAFNLKEGRHLQKGDSKKILIHEELAKKNGLSLHDKIGLDAGQSE-S 182
```

```
Query: 181 GKIKSHKFKIIGIFSGKKQETYTGLSSDFSENMVFVDYSTSQEILNKSENNRIANKILMY 240
           GK ++ +F+IIGIFSGKKQE +TGLSSDFSEN VF DY +SQ +L  SE    A +   Y
Sbjct: 183 GKGQTVEFEIIGIFSGKKQEKFTGLSSDFSENQVFTDYESSQTLLGNSEAQVSAARF--Y 240

Query: 241 SGSLESTELALNKLKDFKIDKSKYSIKKDNKAFEESLESVSGIKHIIKIMTYSIMLGGIV 300
             + +  +  + ++++  ++   Y ++K+NKAFE+  +SV+  +  + I  Y +++ G
Sbjct: 241 VENPKEMDGLMKQVENLALENQGYQVEKENKAFEQIKDSVATFQTFLTIFLYGMLIAGAG 300

Query: 301 VLSLILILWLRERIYEIGIFLSIGTTKIQIIRQFIFELIFISIPSIISSLFLGNLLLKVI 360
            L L+L LWLRER+YE+GI L++G  K  I  QF  E++ +S+ +++ +   GN +   +
Sbjct: 301 ALILVLSLWLRERVYEVGILLALGKGKSSIFLQFCLEVVLVSLGALLPAFVAGNAITTYL 360

Query: 361 VEGFINSENSMIFGGSLINKSSFMLNITTLAESYLILISIIVLSVVMASSLILFKKPKEI 420
           ++  + S +      +L   SS   +I + AESY+ L+ +  LSV +    +  K PKEI
Sbjct: 361 LQTLLASGDQASLQDTLAKASSLSTSILSFAESYVFLVLLSCLSVALCFLFLFRKSPKEI 420

Query: 421 LSKIS 425
           LS IS
Sbjct: 421 LSSIS 425
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1359> which encodes the amino acid sequence <SEQ ID 1360>. Analysis of this protein sequence reveals the following:

```
        Possible site: 15
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-11.57    Transmembrane    23 -  39 (  16 -  43)
   INTEGRAL    Likelihood =-11.36    Transmembrane   371 - 387 ( 362 - 396)
   INTEGRAL    Likelihood = -8.12    Transmembrane   331 - 347 ( 324 - 360)
   INTEGRAL    Likelihood = -7.70    Transmembrane   280 - 296 ( 277 - 308)

----- Final Results -----
               bacterial membrane --- Certainty=0.5628(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAB97962 GB:U96166 ATP-binding cassette transporter-like protein
          [Streptococcus cristatus]
 Identities = 222/311 (71%), Positives = 278/311 (89%)

Query: 16  MRSILTMLGIIIGIGAIIAIFSIIEGNTENTKRQLIGGSNNTINIVFNKKSSIDPKFPDK 75
           MRS+LTMLGIIIGIGAIIAIFSIIEGNTENTKRQLIGGSNNTI +V++KKS+IDP  P+K
Sbjct: 1   MRSMLTMLGIIIGIGAIIAIFSIIEGNTENTKRQLIGGSNNTIKVVYDKKSAIDPSIPEK 60

Query: 76  SNAKKPDYLPFMAEEELSKIQQVKGVKNALISYGIDDKVYHLGQKSSAKISAITKNVAEV 135
           S A+KP Y+PFM E+ LSKI+++ GVKNAL++YG D+K+Y+L QKSS+K+ A++++VA++
Sbjct: 61  SQAQKPSYIPFMGEDVLSKIKEIPGVKNALMTYGADEKIYYLSQKSSSKVQAVSQSVADI 120

Query: 136 RRMTFIKGSDFSDKDFIDQKQVIYLEKSLYESLFPKDDGLGKFVEVMGNPFRVIGVFESK 195
           ++   ++G  F  + F +Q+QV YLEKSLY++LFPK DG+GK+VEV GNPF+VIGVFES
Sbjct: 121 KQQRLLEGEGFDSEAFKNQEQVAYLEKSLYDTLFPKGDGIGKYVEVKGNPFKVIGVFEST 180

Query: 196 EQSGLTSGTEKIAYIPLHQWYNINGVVDATPEITIQTYRADDLKPVAKRVSDMLNQTIPK 255
           EQSGLTSG+EK+AYIPL QW+ I   ++ +PE+T+QT++ADDLK VAK+VSD LNQ +P+
Sbjct: 181 EQSGLTSGSEKVAYIPLQQWHRIFDTINVSPEVTVQTHKADDLKKVAKKVSDYLNQQMPQ 240

Query: 256 SDYMFGVMNLKEFERQLDNLNKSNFVLLAGIASISLIVGGIGVMNIMLVSVTERTREIGI 315
           SDYMFGV+NL+EFERQLDNLN+SNFVLLAGIASISL+VGGIGVMNIMLVSVTERTREIGI
Sbjct: 241 SDYMFGVLNLQEFERQLDNLNQSNFVLLAGIASISLLVGGIGVMNIMLVSVTERTREIGI 300

Query: 316 KKALGARRKLI 326
           KKALGARRK++
Sbjct: 301 KKALGARRKIL 311
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 79/386 (20%), Positives = 170/386 (43%), Gaps = 38/386 (9%)

Query: 5   AIAYITRKKNRTLIIFAILTIVLSCLYSCLTIMKSSNE-IEKALYESSNSSISITKKDGK 63
            A++ I   K R+++    + I +  + +  +I++ + E  ++ L   SN++I+I
Sbjct: 7   ALSSILSHKMRSILTMLGIIIGIGAIIAIFSIIEGNTENTKRQLIGGSNNTINIV----- 61

Query: 64  YFNINQFKNIEKIKEVEEKIFQYDGLAKLKDLKVVSGEQSINREDLSDEFKNVVSLEATS 123
             FN        K  ++  K F     AK  D      E+ +++       KN +
Sbjct: 62  -FN--------KKSSIDPK-FPDKSNAKKPDYLPFMAEEELSKIQQVKGVKNALISYGID 111

Query: 124 NTKRNLLFSSGVFSFKEGKNIEENDKNSILVHEEFAKQNKLKLGDEIDLELLDTE----- 178
            +   +L   S        KN+ E  + + +   +F+ ++ +     I LE    E
Sbjct: 112 DKVYHLGQKSSAKISAITKNVAEVRRMTFIKGSDFSDKDFIDQKQVIYLEKSLYESLFPK 171

Query: 179 -----KSGKIKSHKFKIIGIFSGKKQETYTGLSSDFSENMVFVDYSTSQEILNKSENNRI 233
                 K  ++  + F++IG+F  K+Q   +GL+S  +E + ++       I    +
Sbjct: 172 DDGLGKFVEVMGNPFRVIGVFESKEQ---SGLTSG-TEKIAYIPLHQWYNINGVVDATPE 227

Query: 234 ANKILMYSGSLESTELALNKLKDFKIDKSKYSIKKDN-KAFEESLESVSGIKHIIK--IM 290
                   +  L+     ++ + +  I KS Y     N K FE  L++++    ++   I
Sbjct: 228 ITIQTYRADDLKPVAKRVSDMLNQTIPKSDYMFGVMNLKEFERQLDNLNKSNFVLLAGIA 287

Query: 291 TYSIMLGGIVVLSLILILWLRERIYEIGIFLSIGTTKIQIIRQFIFELIFIS----IPSI 346
            + S+++GGI V++++L+  + ER  EIGI  ++G  +  I++QF+ E + ++    +  +
Sbjct: 288 SISLIVGGIGVMNIMLVS-VTERTREIGIKKALGARRKLILKQFLIEAVILTLLGGVIGV 346

Query: 347 ISSLFLGNLLLKVIVEGFINSENSMI 372
            IS +  G ++ + +   +I S  S++
Sbjct: 347 ISGMVSGLIITRSLEYPYILSLFSVV 372
```

A related GBS gene <SEQ ID 8571> and protein <SEQ ID 8572> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop Possible site: -1   Crend: 10
McG: Discrim Score:      5.59
GvH: Signal Score (-7.5): -5.97
     Possible site: 56
>>> Seems to have an uncleavable N-term signal seq
ALOM program   count: 4 value: -11.52 threshold:  0.0
   INTEGRAL    Likelihood =-11.52   Transmembrane  294 - 310 ( 285 - 312)
   INTEGRAL    Likelihood =-10.67   Transmembrane  396 - 412 ( 385 - 417)
   INTEGRAL    Likelihood = -8.76   Transmembrane   17 -  33 (  14 -  38)
   INTEGRAL    Likelihood = -4.14   Transmembrane  335 - 351 ( 333 - 357)
   PERIPHERAL  Likelihood =  4.51     315
 modified ALOM score:   2.80

*** Reasoning Step: 3

----- Final Results -----
              bacterial membrane --- Certainty=0.5607(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
38.7/67.3% over 421aa
                                                                   Streptococcus
pneumoniae
  GP|5712667| Vexp1 Insert characterized

ORF00815(304 - 1575 of 1875)
GP|5712667|gb|AAD47592.1|AF140784_1|AF140784(4  -  425  of  425)  Vexp1  {Streptococcus
pneumoniae}
%Match = 25.0
%Identity = 38.7  %Similarity = 67.2
Matches = 164  Mismatches = 136  Conservative Sub.s = 121
```

```
48        78        108       138       168       198       228       258
SIEH*VVFDNKTI*T*ELDFVSHSS**VI*DFPLNK*IRNSVTSYINGSIIEIVCQMKKF*WK*F*KH*L*AM*KY*SSG

288       318       348       378       408       438       468       495
CNSCGVKIERSN*EVIKNAIAYITRKKNRTLIIFAILTIVLSCLYSCLTIMKSSNEIEKALYESSNSSISITK-KDGKYF
               |: : ||::||: |::|:| || ::|: : :|||:|||:  :|  ||:| |:| || | ::|: |
            MNPIQRSWAYVSRKRLRSFILFLILLVLLAGISACLTLMKSNKTVESNLYKSLNTSFSIKKIENGQTF
                    10        20        30        40        50        60

525       555       585       615       645       675       705       735
NINQFKNIEKIKEVEEKIFQYDGLAKLKDLKVVSGEQSINREDLSDEFKNVVSLEATSNTKRNLLFSSGVFSFKEGKNIE
 :: : :: ||| :|    : : :||||| : |:||||: |:|||    |:||| |  :: ::: |:|  |::|||::::
KLSDLASVSKIKGLENVSPELETVAKLKDKEAVTGEQSVERDDLSAADNNLVSLTALEDSSKDVTFTSSAFNLKEGRHLQ
          80        90       100       110       120       130       140

765       795       825       855       885       915       945       975
ENDKNSILVHEEFAKQNKLKLGDEIDLELLDTEKSGKIKSHKFKIIGIFSGKKQETYTGLSSDFSENMVFVDYSTSQEIL
: |   ||:|||:||:| | | |:| |:   :| ||| :: :|:|||||||||| :|||||||||| || || :|| :|
KGDSKKILIHEELAKKNGLSLHDKIGLDAGQSE-SGKGQTVEFEIIGIFSGKKQEKFTGLSSDFSENQVFTDYESSQTLL
         160       170       180       190       200       210       220

1005      1035      1065      1095      1125      1155      1185      1215
NKSENNRIANKILMYSGSLESTELALNKLKDFKIDKSKYSIKKDNKAFEESLESVSGIKHIIKIMTYSIMLGGIVVLSLI
  ||     :    |  : :  :  : ::::: :: ::   | ::|:|||||:  :||:  :  : |  | ::: |   | |:
GNSEA--QVSAARFYVENPKEMDGLMKQVENLALENQGYQVEKENKAFEQIKDSVATFQTFLTIFLYGMLIAGAGALILV
             240       250       260       270       280       290       300

1245      1275      1305      1335      1365      1395      1425      1455
LILWLRERIYEIGIFLSIGTTKIQIIRQFIFELIFISIPSIISSLFLGNLLLKVIVEGFINSENSMIFGGSLINKSSFML
| ||||||:||:||:|::|  |  |  || :|::::|: ::: ::  || :   ::: :: | :   :  :|   ||:  
LSLWLRERVYEVGILLALGKGKSSIFLQFCLEVVLVSLGALLPAFVAGNAITTYLLQTLLASGDQASLQDTLAKASSLST
             320       330       340       350       360       370       380

1485      1515      1545      1575      1605      1635      1665      1695
NITTLAESYLILISIIVLSVVMASSLILFKKPKEILSKIS*EQIMDILEIKNVNYSYANSKEKVLSGVNQKFELGKFYAI
:| ::||||: |: :  ||| :   ::: | |||||| ||
SILSFAESYVFLVLLSCLSVALCFLFLFRKSPKEILSSIS
             400       410       420
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 419

A DNA sequence (GBSx0455) was identified in *S.agalactiae* <SEQ ID 1361> which encodes the amino acid sequence <SEQ ID 1362>. Analysis of this protein sequence reveals the following:

```
Possible site: 42
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -5.04   Transmembrane   19 -  35 (  14 -  42)

----- Final Results -----
               bacterial membrane --- Certainty=0.3017(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 420

A DNA sequence (GBSx0456) was identified in *S.agalactiae* <SEQ ID 1363> which encodes the amino acid sequence <SEQ ID 1364>. Analysis of this protein sequence reveals the following:

```
Possible site: 26
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 421

A DNA sequence (GBSx0457) was identified in *S.agalactiae* <SEQ ID 1365> which encodes the amino acid sequence <SEQ ID 1366>. Analysis of this protein sequence reveals the following:

```
Possible site: 17
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
                bacterial outside --- Certainty=0.3000(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAA74029 GB:U30715 ORFB [Bacillus anthracis]
 Identities = 33/76 (43%), Positives = 44/76 (57%), Gaps = 1/76 (1%)

Query: 11  IRRVSHACTKAGDRFYEENILNREFTATAHNQKWCTDVTYLQYGLGAKAYLSAIKDLYNG 70
            ++R          R   EN+LNR F A   N+KW TD+TYL +G     YL +I DLYN
Sbjct: 86  VKRKRRTWINGESRIVVENLLNRNFQANKPNEKWVTDITYLPFGT-EMLYLLSIMDLYNN 144

Query: 71  SIIAYEISHNNEIHLL 86
             IIAYEIS+  ++ L+
Sbjct: 145 EIIAYEISNRQDVTLV 160
```

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 422

A DNA sequence (GBSx0458) was identified in *S.agalactiae* <SEQ ID 1367> which encodes the amino acid sequence <SEQ ID 1368>. Analysis of this protein sequence reveals the following:

```
Possible site: 27
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.69   Transmembrane   10 -  26 (  10 -  26)

----- Final Results -----
               bacterial membrane --- Certainty=0.1277(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 423

A DNA sequence (GBSx0459) was identified in *S.agalactiae* <SEQ ID 1369> which encodes the amino acid sequence <SEQ ID 1370>. Analysis of this protein sequence reveals the following:

```
Possible site: 47
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4170(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAA56999 GB:U09558 ORFA, putative Helix-Turn-Helix motif from
           amino acid 21 through 42 and from amino acid 78 through
           99 [Lactobacillus johnsonii]
 Identities = 28/116 (24%), Positives = 59/116 (50%), Gaps = 6/116 (5%)

Query: 3    YSTLAKEQGVQGYLDGKGSLRDICKWYDISSRSVLQKWIKRYTSGEDLKATSRGYSRMKQ 62
            YST  K + V  YL+ + S++ + K Y+I   +++++W+ +    + L A S   +++
Sbjct: 4    YSTELKIEIVSKYLNHEDSIKGLAKQYNIHW-TLIRRWVDK-AKCQGLAALSVKHTKTTY 61

Query: 63   GRQATFEERVEIVNYTIAHGKDYQAAIEKFGVSYQQIYSWVRKLEKNGSQGLVDRR 118
                  + ++ +V Y + H         KF +S  Q+Y+W +K  + G  GL+ ++
Sbjct: 62   SS----DFKLNVVRYYLTHSIGVSKVAAKFNISDSQVYNWAKKFNEEGYAGLLPKQ 113
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 424

A DNA sequence (GBSx0460) was identified in *S.agalactiae* <SEQ ID 1371> which encodes the amino acid sequence <SEQ ID 1372>. Analysis of this protein sequence reveals the following:

```
Possible site: 23
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -0.69    Transmembrane     2 -  18 (    2 -  19)

----- Final Results -----
              bacterial membrane --- Certainty=0.1277(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 425

A DNA sequence (GBSx0461) was identified in *S.agalactiae* <SEQ ID 1373> which encodes the amino acid sequence <SEQ ID 1374>. This protein is predicted to be integrase (phage-relatedpr). Analysis of this protein sequence reveals the following:

```
Possible site: 28
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAC79517 GB:U88974 ORF1 [Streptococcus thermophilus temperate
           bacteriophage 01205]
 Identities = 104/172 (60%), Positives = 127/172 (73%), Gaps = 11/172 (6%)

Query: 10  QHQSYAALYLIAKTGMRFAECLGLTVNDIDYTNKYLSINKTWDYHFNQRYLPTKNKSSIR 69
            ++ SYAALY+I+KTG+RFAECLGLTV+DI      LS+NKTWDY  N  ++PTK KSSIR
Sbjct: 186 EYASYAALYIISKTGIRFAECLGLTVDDIKRDTGMLSVNKTWDYKNNTGFMPTKTKSSIR 245

Query: 70  NIPIDNDTLFFLHEFTKNKNDRLFDKLSNNAVNKTIRKITGREVRVHSLRHTFASY---- 125
             IP+D++ + F+ +     + RL   LSNNAVNKT+RKI GREVRVHSLRHT+ASY
Sbjct: 246 EIPLDDEFINFIDQLPPTDDGRLLPSLSNNAVNKTLRKIVGREVRVHSLRHTYASYLIAH 305

Query: 126 ---LISISQVLDHENLNITLEVYAHQLQEQKDRNDKLNQRNLGRIWGKIALN 174
               LIS+SQVL HENLNITLEVYAHQLQEQK RND+     + ++W K   N
Sbjct: 306 DIDLISVSQVLGHENLNITLEVYAHQLQEQKSRNDE----KIKQMWTKCGQN 353
```

There is also homology to SEQ ID 578

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 426

A DNA sequence (GBSx0462) was identified in *S.agalactiae* <SEQ ID 1375> which encodes the amino acid sequence <SEQ ID 1376>. Analysis of this protein sequence reveals the following:

```
Possible site: 22
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3206(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database, but there is homology to SEQ ID 1328.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 427

A DNA sequence (GBSx0463) was identified in *S.agalactiae* <SEQ ID 1377> which encodes the amino acid sequence <SEQ ID 1378>. Analysis of this protein sequence reveals the following:

```
Possible site: 45
```

```
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.6542(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB52541 GB:AJ131519 hypothetical protein [Lactobacillus
          bacteriophage phi adh]
 Identities = 24/55 (43%), Positives = 36/55 (64%)

Query: 12 MDKELTPQEKANKKWAENNREHRTYLSKRSTARSFINKNATKEDLLELKQLIESK 66
           M K    + KANKKW E N+  + Y++KRSTA+SFI   AT+EDL  +++ +  +
Sbjct: 1  MAKITEARAKANKKWDEKNKARKLYINKRSTAKSFILNLATEEDLANIEEYVAER 55
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 428

A DNA sequence (GBSx0464) was identified in *S.agalactiae* <SEQ ID 1379> which encodes the amino acid sequence <SEQ ID 1380>. Analysis of this protein sequence reveals the following:

```
Possible site: 41
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.4417(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database, but there is homology to SEQ ID 1332.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 429

A DNA sequence (GBSx0465) was identified in *S.agalactiae* <SEQ ID 1381> which encodes the amino acid sequence <SEQ ID 1382>. Analysis of this protein sequence reveals the following:

```
Possible site: 28
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
                 bacterial outside --- Certainty=0.3000(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 430

A DNA sequence (GBSx0466) was identified in *S.agalactiae* <SEQ ID 1383> which encodes the amino acid sequence <SEQ ID 1384>. Analysis of this protein sequence reveals the following:

```
Possible site: 47
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -4.30   Transmembrane  205 - 221 ( 202 - 223)
   INTEGRAL    Likelihood = -3.56   Transmembrane  296 - 312 ( 294 - 312)

----- Final Results -----
              bacterial membrane --- Certainty=0.2720(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9663> which encodes amino acid sequence <SEQ ID 9664> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

A related GBS gene <SEQ ID 8573> and protein <SEQ ID 8574> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 8
McG: Discrim Score:     -8.80
GvH: Signal Score (-7.5): -4.03
     Possible site: 47
>>> Seems to have no N-terminal signal sequence
ALOM program   count: 2 value:  -4.30 threshold:  0.0
   INTEGRAL    Likelihood = -4.30   Transmembrane  205 - 221 ( 202 - 223)
   INTEGRAL    Likelihood = -3.56   Transmembrane  296 - 312 ( 294 - 312)
   PERIPHERAL  Likelihood =  2.97      20
 modified ALOM score:   1.36

*** Reasoning Step: 3

----- Final Results -----
              bacterial membrane --- Certainty=0.2720(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

SEQ ID 8574 (GBS366) was expressed in *E.coli* as a GST-fusion product. The purified fusion protein (Figure 215, lane 5) was used to immunise mice. The resulting antiserum was used for FACS (Figure 281), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 431

A DNA sequence (GBSx0467) was identified in *S.agalactiae* <SEQ ID 1385> which encodes the amino acid sequence <SEQ ID 1386>. This protein is predicted to be N-acetylmuramoyl-L-alanine amidase. Analysis of this protein sequence reveals the following:

```
Possible site: 31
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.1471(Affirmative) < succ>
```

```
                    bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                     bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 8575> which encodes amino acid sequence <SEQ ID 8576> was also identified. This has an RGD motif at residues 81-83.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB07986 GB:Z93946 N-acetylmuramoyl-L-alanine amidase
          [bacteriophage Dp-1]
 Identities = 99/140 (70%), Positives = 120/140 (85%)

Query: 10   MVINIEQAIAWMASRKGKVTYSMDYRNGPSSYDCSSSVYFALRSAGASDNGWAVNTEYEH 69
            M ++IE+ +AWM +RKG+V+YSMD+R+GP SYDCSSS+Y+ALRSAGAS  GWAVNTEY H
Sbjct: 1    MGVDIEKGVAWMQARKGRVSYSMDFRDGPDSYDCSSSMYYALRSAGASSAGWAVNTEYMH 60

Query: 70   DWLIKNGYVLIAENTNWNAQRGDIFIWGKRGASAGAFGHTGMFVDPDNIIHCNYGYNSIT 129
             WLI+NGY LI+EN  W+A+RGDIFIWG++GASAGA GHTGMF+D DNIIHCNY Y+ I+
Sbjct: 61   AWLIENGYELISENAPWDAKRGDIFIWGRKGASAGAGGHTGMFIDSDNIIHCNYAYDGIS 120

Query: 130 VNNHDEIWGYNGQPYVYAYR 149
            VN+HDE W Y GQPY Y YR
Sbjct: 121 VNDHDERWYYAGQPYYYVYR 140
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1387> which encodes the amino acid sequence <SEQ ID 1388>. Analysis of this protein sequence reveals the following:

```
     Possible site: 26

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.06    Transmembrane    79 -  95 (  77 -  95)

----- Final Results -----
                    bacterial membrane --- Certainty=0.1426(Affirmative) < succ>
                     bacterial outside --- Certainty=0.0000(Not Clear) < succ>
                   bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 56/91 (61%), Positives = 68/91 (74%)

Query: 158 KVDNQSVVSKFEKELDVNTPLSNSNMPYYEATISEDYYVESKPDVNSTDKELLVAGTRVR 217
            K+D        F ++LD NT L NSN+PYYEAT+  DYYVESKP+ +S DKE + AGTRVR
Sbjct: 354 KIDKPQSQLTFNQKLDTNTKLDNSNVPYYEATLRTDYYVESKPNASSADKEFIKAGTRVR 413

Query: 218 VYEKVKGWARIGAPQSNQWVEDAYLIDATDM 248
            VYEKV GW+RI A QS+QWVED YL +AT +
Sbjct: 414 VYEKVNGWSRINASQSDQWVEDKYLSNATQV 444
```

SEQ ID 8576 (GBS301) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 44 (lane 9; MW 30kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 49 (lane 3; MW 55kDa).

The GBS301-GST fusion product was purified (Figure 205, lane 4) and used to immunise mice. The resulting antiserum was used for FACS (Figure 300), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 432

A DNA sequence (GBSx0468) was identified in *S.agalactiae* <SEQ ID 1389> which encodes the amino acid sequence <SEQ ID 1390>. Analysis of this protein sequence reveals the following:

```
Possible site: 53
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -6.53   Transmembrane    8 -  24 (   3 -  25)

----- Final Results -----
               bacterial membrane --- Certainty=0.3612(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 433

A DNA sequence (GBSx0469) was identified in *S.agalactiae* <SEQ ID 1391> which encodes the amino acid sequence <SEQ ID 1392>. Analysis of this protein sequence reveals the following:

```
Possible site: 34
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
                bacterial outside --- Certainty=0.3000(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 434

A DNA sequence (GBSx0470) was identified in *S.agalactiae* <SEQ ID 1393> which encodes the amino acid sequence <SEQ ID 1394>. Analysis of this protein sequence reveals the following:

```
Possible site: 36
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0120(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 435

A DNA sequence (GBSx0471) was identified in *S.agalactiae* <SEQ ID 1395> which encodes the amino acid sequence <SEQ ID 1396>. Analysis of this protein sequence reveals the following:

```
Possible site: 38
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4757(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9661> which encodes amino acid sequence <SEQ ID 9662> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 436

A DNA sequence (GBSx0472) was identified in *S.agalactiae* <SEQ ID 1397> which encodes the amino acid sequence <SEQ ID 1398>. This protein is predicted to be a minor structural protein. Analysis of this protein sequence reveals the following:

```
Possible site: 23
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -2.39    Transmembrane  349 - 365 ( 347 - 366)

----- Final Results -----
               bacterial membrane --- Certainty=0.1956(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAF43531 GB:AF145054 ORF39 [Streptococcus thermophilus
           bacteriophage 7201]
 Identities = 212/666 (31%), Positives = 323/666 (47%), Gaps = 52/666 (7%)

Query: 10  WGNNLTLEILSAWNKP---NIASNTSTVNVQVFL-----KMSSYGYISIGETRPLKITVD 61
            W NN     +  W      +I +NTS V +++ L       + Y   +  E    ++
Sbjct: 5   WSNNDRGYRIRLWVDQVGQDIQNNTSQVRLRLSLLNTTTTFAQYSCSAFVEFNGQRLNWS 64

Query: 62  GRAETINVNPSINYGQRKLLFAKDYIVNHNSDGNKPLFNISAYYPIN--FSNYGEATANQ 119
            G    +  N +I       L  +   V H  DG+  +F + A++  +  +S       NQ
Sbjct: 65  GSPSVLGWNQTIQ------LIDQTITVRHADDGSG-VFGVHAHFNGSGGWSPGNLDIGNQ 117

Query: 120 SISLPKINRLSVSSAISGVLGNAVTITINRYSTSFTHNLKYDFKGSTGTIATGVGTSYLW 179
             I+L  I R S      G +GN V I+I+R      TH L+Y ++  G IA  VGTSY W
Sbjct: 118 QITLTTIPRGSSVRVSDGFIGNQVDISIDRKIGGATHTLRYAWENKQGKIADNVGTSYKW 177

Query: 180 TIPPTFANLLPNELTGTGNLIVETMDGSAKIGETKYTLSITIPNTATYKPKLSSITLSDT 239
            TIP  FAN +PN  +G G + V+T      I     TL+ ++  T   KP  +  TL+DT
Sbjct: 178 TIPEDFANDIPNSTSGRGTIYVDTYINGNFIQTQSTTLTASV-ITNNLKPSFTGFTLTDT 236

Query: 240 NTLTSSIVSG-NNFVRIISKVKVDFGSAIGNNGSTITSYNAEIVGKSNSIIGNGSVFDKL 298
            N  +  IV G  +FV I+S VKV F  A   +G+TI  Y AEIVG +NSI  NG V  ++
Sbjct: 237 NPTSQRIVPGQTHFVSIMSLVKVVFNGAQAKSGATIVGYYAEIVGANNSISSNGGVLREV 296
```

```
Query: 299 DFFGSA--TIRATVTDSRGLTSEPVDTKINVIDYFLPIVTSAKVVRSQQNPDILQVLPFV 356
                   T+R  V DSRG+ S+ V+TK+  + YF P +   +V RS +  DIL +  F
Sbjct: 297 SVNQDTEMTLRGRVQDSRGIWSDWVETKLTFLFYFSPAL-RFEVKRSDKKLDILTIKRFA 355

Query: 357 KIAPIIVGGIQKNQLKMSVSVAPYNTGIYAVDSGAATNTWSTISQMSGAPLNLGGTYDKS 416
           KIAP+ V GIQ+N +K++ S A      + VD+G A   WS+IS+ + +   LG +Y
Sbjct: 356 KIAPLSVNGIQRNVMKLTFSTAKVGWDNFVVDNGQAGGVWSSISEFNASDAKLGNSYPAD 415

Query: 417 KSWLVKISVSDNLMSATPIIQPVASEFVLVTKAPSGVAFGKIWEHGIIDAKGDVYVDGTI 476
            S++V   + D   S T     V ++ V++T    GV  GK  E G +D  GD      I
Sbjct: 416 TSYVVIGKLEDEFTS-TSFQATVPTDEVIMTYDRQGVGIGKYRERGALDVNGD------I 468

Query: 477 YCGDKAIQQKPLALNNGGSFRHDDTDLNSLQDTGFYCVFRGANRPAGAGPGYVTVVRHET 536
           Y  +  IQQ  L  NNG     ++   N+++D G Y +F  A  P      +  +  H +
Sbjct: 469 YANNSPIQQYQLTNNNGSPKMTNNA--NTIEDPGQYYLFSAA--PGNPSGQWGHLFHHSS 524

Query: 537 ------ANYAYQQFYDRTNKTI-----FTRLLENGVWSGWSEYVKKD--SLQTTGWITIG 583
                    A Q F+    +       ++R++++  W  W E+ + D  +L  TGW   G
Sbjct: 525 YGKGSMYKEAIQIFWSNDGRLFSRHHRWSRIIDD--WEPWKEFARNDNTNLINTGWQPAG 582

Query: 584 -NGFKYKRKGDDIDLMYNFASNGLQRWSVGNMPSGLI--PQELMFAITGWTLAPDKSIHL 640
            +G YKR GD + + +NF   G   + + ++P  +   PQ  MF +TGW++  +K  ++
Sbjct: 583 VDGSFYKRVGDVLTIKFNFTGTG-GDFLLASVPPEIFKAPQSYMFVVTGWSVWANKQYNV 641

Query: 641 QINASG 646
           Q+N  G
Sbjct: 642 QVNEGG 647
```

No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 1398 (GBS365) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 82 (lane 2; MW 102kDa).

GBS365-GST was purified as shown in Figure 216, lane 11.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 437

A DNA sequence (GBSx0473) was identified in *S.agalactiae* <SEQ ID 1399> which encodes the amino acid sequence <SEQ ID 1400>. This protein is predicted to be a minor structural protein. Analysis of this protein sequence reveals the following:

```
Possible site: 59
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3481(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAC34413 GB:AF158600 putative minor structural protein
          [Streptococcus thermophilus bacteriophage Sfi11]
 Identities = 504/998 (50%), Positives = 675/998 (67%), Gaps = 56/998 (5%)

Query: 1   MLTIHGPDLKPVLFLDNDKQGALNYFNHKWYRKQKTGSSVLEFSVYKKDLLGDSPLSHKY 60
           +LTIH  +L+ V ++DN+KQ  LN+FN KW R  ++G+SV EFSV+KK +  DS +    Y
Sbjct: 2   LLTIHDNNLQKVAYIDNEKQSTLNFFNDKWTRSLESGTSVFEFSVFKKSIKSDSKVEISY 61

Query: 61  HVLNDQAFVSFVHKGKVQLLNIMKIDEDEKQIDCYCENLNLELLNEYCNAYKATKAMSFE 120
```

```
             LN++AFVSF HKGK  L N+MKI+EDE+ I CYCENL+LELL EY  AYKA+K M+F+
Sbjct: 62  KYLNERAFVSFKHKGKSYLFNVMKIEEDEQIIRCYCENLSLELLLEYRGAYKASKPMTFK 121

Query: 121 EYLVQFDILSWGALTVGTNEVKDKKLTLEWTSQETKLARLLSIANNFDAEIEFETKLNFN 180
           EY   + +  +  LT+G NEV D+K TLEW  QET LARL+S+A NFDAEIEF+T+L  N
Sbjct: 122 EYFDDWGMGQFAKLTLGVNEVSDQKRTLEWEGQETTLARLISLARNFDAEIEFDTRLKPN 181

Query: 181 HTFKQLIINIYKEYEEGKSYGVDRDKTDVILRYQKNISGIRKTVDKRQIYNAIRPYGKK- 239
               + ++N+YK Y+ GK+ GV R ++DVIL+Y KNI+GI+++VDK QIYN I PYG+K
Sbjct: 182 SQLDEFVLNVYKAYD-GKNQGVGRRRSDVILKYGKNINGIKRSVDKTQIYNMITPYGRKS 240

Query: 240 -TVRGERVISNPVTRKVTKTVGSNRT---YLGGDLKYYGHTIKKANVQAIINYAVQYNIL 295
            T +  + IS+PVT +    V S R    Y GGDL Y GHT+  + VQ I N  VQ N+L
Sbjct: 241 DTKKETKRISDPVTIQNPVVVPSARVEKRYAGGDLTYAGHTLSASLVQTIFNLCVQRNLL 300

Query: 296 PSGIITQLYLESFWGDSTVGKRDNNWAGMSGGAQTRPSGVKVTTGMARPANEGGTYMHYA 355
           PSG+I+QLYLESFWG S V +RDNNW+GM+GGAQTRPSGV VTTG  RPA+EGGTYMHYA
Sbjct: 301 PSGVISQLYLESFWGSSNVARRDNNWSGMTGGAQTRPSGVVVTTGSPRPASEGGTYMHYA 360

Query: 356 SVDDFLKDYTYLLAKQG-----IYNVVGKKNIADYTKGLFRAGGAKYDYAAAGYQSYTNL 410
           SVDDF+KDYTYLLA Q      +Y V GK+NI +YTKGLFR GGA YDYAAAGY  Y  L
Sbjct: 361 SVDDFMKDYTYLLADQTSGGRKMYGVKGKQNIEEYTKGLFRIGGALYDYAAAGYNHYIYL 420

Query: 411 MTNIRNGINKVTGNILNTIDKLWQTPVKPITAVNVARRATKTIQA------INEATKLKG 464
           M +IRNGIN+  GNIL+ +D LW+ P  IT  N  ++ T+T++A      +NE   LKG
Sbjct: 421 MRDIRNGINRSNGNILDKLDDLWRQPDNQITQPN--KQVTRTVKADRVIAVLNEMQGLKG 478

Query: 465 RRIGSGQCYALSGWYAKKLDGAWIDSSIGGIRGRIGGGMAAALIGTDYNWGAYGWKVDKS 524
           RR+G+GQCYAL+ WY+ KL G  + + + G  G IG GMAAA IGTDY W  +GW V +
Sbjct: 479 RRVGNGQCYALAAWYSMKLGGPGLGAGVTGKSGVIGAGMAAAKIGTDYAWDRFGWSVVRP 538

Query: 525 PNAGNLKAGGIYNVRANRGAPFYTTGWGHTGIIKSVSKTRVTVLEQNFVGRMYVVENSYD 584
            +   LK G I N++A       T+ WGH  II S + + VTVLEQN+ GR YVV+NSY
Sbjct: 539 TSVDQLKPGAIANIKAYNSY-LGTSVWGHVSIIISNNGSTVTVLEQNYAGRQYVVQNSYP 597

Query: 585 INSFASGLQTVCYPREIAQGMSVNGATTQQVSGGTQISYEEVVQEAQTESYEEEQIIYID 644
            +++   ++T+CYP E+ +G +V G T    +   ++   E+       + E    + ID
Sbjct: 598 ASAYLGAVETLCYPPELKEGKTVEGRTETVSTPNVEVQKVEIPPIDVEVTTESTAALTID 657

Query: 645 NSIYKEWKDENGKVEYYLKNGFLYAPLSRDRYPSVLTGNETRDNWIRKDMEVETDSQEVL 704
           +   +EW++ENG+VE+YL+NG LYAP+S++ YPS+LTG E  DNWIRKDME++TDS++VL
Sbjct: 658 SKRKQEWRNENGQVEFYLENGSLYAPISKELYPSILTGKENGDNWIRKDMEIDTDSEDVL 717

Query: 705 MSTGLKDLKAHAYPAITYEVDGYVDLELGDVVRIQDDGYEPPLILTARVVEQEISITNPS 764
           +ST L++L+   YPAITYEVDG++DL++GD V+IQD G+ P L+L ARV EQ+IS TNP
Sbjct: 718 ISTALRNLRKFCYPAITYEVDGFLDLDIGDTVKIQDTGFSPMLMLEARVSEQQISFTNPV 777

Query: 765 SNKTKFSNFVEKESQLASDLISDMLRLYDESIPYEIKLATSNGVAFKNGTGESVLTPSLQ 824
            NKT F+NF   +++++  L+S M +L +E+IPYE+KL+T NG  FKN TG+SVL  +L+
Sbjct: 778 ENKTVFANFQTLQNKVSDSLLSRMTKLAEEAIPYELKLSTDNGTTFKNSTGQSVLKATLE 837

Query: 825 KNGKDYEAVYFYKNGDSLIDIGPSLIVKASDFNHVLNITVEAYLNEELVASTQISFTDTE 884
           KNG+ Y+ ++F+KNGDS+I  G  L+VK +DF + L +TVEAYL++ELVAS +I+FTD
Sbjct: 838 KNGEVYQPIFFFKNGDSIIGTGNQLVVKPTDFENTLQVTVEAYLDDELVASAEITFTDVS 897

Query: 885 DGADGKDGAPGPQGPPGVNGLQGPKGDQGIQGPAGADGKATYTHIAYALDENGSTGFSVS 944
           DG                   QGPKGD G+                  L    S G+
Sbjct: 898 DGK------------------QGPKGDDGVS--------------PINLIIESSNGYQFK 925

Query: 945 DNVGKTYI--GMYVDDNIIDSNDPK-KYKWNLIKGADG 979
           +N+  T     +Y D+  ID +  +  Y W+ +  ADG
Sbjct: 926 NNIINTTFTAKLYQDNKEIDKDGTRYAYLWSKV-NADG 962
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1401> which encodes the amino acid sequence <SEQ ID 1402>. Analysis of this protein sequence reveals the following:

```
Possible site: 37
```

```
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -3.56   Transmembrane  325 - 341 ( 323 - 343)

----- Final Results -----
               bacterial membrane --- Certainty=0.2423(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
Identities = 23/55 (41%), Positives = 27/55 (48%)

Query: 886 GADGKDGAPGPQGPPGVNGLQGPKGDQGIQGPAGADGKATYTHIAYALDENGSTG 940
           G  GKDGAPG  G PG  G +G +G+ G QGP G  G+   T         G  G
Sbjct: 181 GEAGKDGAPGKDGAPGEKGEKGDRGETGAQGPVGPQGEKGETGAQGPAGPQGEAG 235
 Identities = 48/151 (31%), Positives = 58/151 (37%), Gaps = 19/151 (12%)

Query: 852 KASDFNHVLNITVEAYLNE--ELVASTQISFTDTEDGADGKDGAPGPQGPPGVNGLQGPK 909
           K  DF   L    E  L E  +L+  + I     + G  G  G  GPQG  G  G QGPK
Sbjct: 82  KEEDFQKELKDFTEKRLKEILDLIGKSGIK---GDRGETGPAGPAGPQGKTGERGAQGPK 138

Query: 910 GD---QGIQGPAGADGKATYTHIAYALDENGSTGFS----VSDNVGKTYIGMYVDDNIID 962
           GD   QGIQG AG  G+           E G  G +         GK       D
Sbjct: 139 GDRGEQGIQGKAGEKGERGEKGDKGETGERGEKGEAGIQGPQGEAGK-------DGAPGK 191

Query: 963 SNDPKKYKWNLIKGADGARGIQGPAGADGKT 993
              P +      +G  GA+G  GP G  G+T
Sbjct: 192 DGAPGEKGEKGDRGETGAQGPVGPQGEKGET 222
 Identities = 25/50 (50%), Positives = 29/50 (58%), Gaps = 9/50 (18%)

Query: 884 EDGADGKDGAPGPQGPPGVNGL---------QGPKGDQGIQGPAGADGKA 924
           +DGA GKDGAPG +G  G  G          QG KG+ G QGPAG  G+A
Sbjct: 185 KDGAPGKDGAPGEKGEKGDRGETGAQGPVGPQGEKGETGAQGPAGPQGEA 234
```

SEQ ID 1400 was expressed in four different forms. SDS-PAGE analysis of total cell extract is shown in Figure 122 (GBS105dN – lane 5 & 7; MW 102kDa), Figure 122 (GBS105dC – lane 8-10; MW 81kDa), Figure 179 (GBS105d – lane 8; MW 102kDa) and in Figure 181 (GBS105C – lane 2; MW 56kDa). GBS105dN-His was purified as shown in Figure 232 (lanes 9 & 10). GBS105dC-His was purified as shown in Figure 233 (lanes 3 & 4).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 438

A DNA sequence (GBSx0474) was identified in *S.agalactiae* <SEQ ID 1403> which encodes the amino acid sequence <SEQ ID 1404>. This protein is predicted to be a minor structural protein. Analysis of this protein sequence reveals the following:

```
Possible site: 32
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2502(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAC34412 GB:AF158600 putative minor structural protein
          [Streptococcus thermophilus bacteriophage Sfi11]
 Identities = 163/433 (37%), Positives = 244/433 (55%), Gaps = 21/433 (4%)
```

```
Query: 80  LSSKKPKMLMFSHIPGRYYLAVQVGDLNFKEIKMNGFGEIT--FIVADAYAHSTSYRRIK 137
           L +KK   L     P RYYLA+  G+++ K I  + + E T  F+V D  AHST+Y+R+
Sbjct: 93  LHTKKAVKLFLPTEPERYYLALVKGEVSLKGIS-DWYDEATIEFLVPDGVAHSTTYKRVT 151

Query: 138 DYTQDGNKMTFKIKNNGTAPAFPIFRIKHLGENGYIGITNETGAFAVGSPEEEDGTIVHR 197
           DY +   KM F I N G+  A+PI  +K   ENGY G+ ++  AF  G+ EE DG I+ +
Sbjct: 152 DYQEKDGKMIFSIDNEGSTDAYPIITLKANAENGYYGLVSDKFAFEAGNIEEADGKIISK 211

Query: 198 NETLFDY-SKAIAQAL-EGAPNVAKLNYMPPTFDSELKRMRLDNILGSGKGGEYVAIGAR 255
            E L+D+    I QA  +GA NV   N       +    + + N+ G         IG +
Sbjct: 212 AEVLYDFRDDRIPQAFAKGAKNVGITNVTGDLHGT----LEIQNVWGRPH------IGLK 261

Query: 256 GTTPGYGE-HVGTRTFIINPDSNGEY-TLNEHLWWKQIFIATAQDQKGFLKLCVTGENDE 313
                  +    + T  I PDS+G    LNE++WW+QIF A +  Q GFLKL V+  +
Sbjct: 262 NPNANINQLQTASLTLDIPPDSSGNVGALNEYIWWRQIFWAGSISQYGFLKLTVSDADGN 321

Query: 314 FLYGIETYKRKNGFETEYNFFALDDDGVGWRFYKQFEFQA-DRNYHNPFSMNRSRAVEIF 372
           FLYG+ET+KR  G E+EYN  A D  G G+RF KQ+ F A +   HNPF+  R  + +I
Sbjct: 322 FLYGVETFKRSLGLESEYNALASDGYG-GFRFLKQWSFLATEYEDHNPFNEPRGWS-DIK 379

Query: 373 REEDKFRIYFNGAHHHVTVPSLKGKKSRKIHLAMGTCSDSSKYINYNLFEKVNFEKMGVS 432
           RE+DK   Y+ G ++  T+P +KGKKS KIHL +     S  ++ +  F+++ + K   +
Sbjct: 380 REDDKVTFYWWGTYNTFTIPEIKGKKSAKIHLTISNI-PSKSFVTHAYFDQLLYIKTNNA 438

Query: 433 HYNNIVNKYQPGDEVIINFENDTVSTKDIDSIQDVVLGSKMISIPPGESELVVHLSSWVA 492
            + +I N+Y  G  +IIN E+DT++  ++ ++ ++V GS    IPPGES++ V  S W
Sbjct: 439 FFEDIPNRYIQGSNLIINSEDDTLTLNNLLNLDEIVDGSLWPVIPPGESQIEVVQSPWAK 498

Query: 493 ALPDISIDFEERY 505
             P ++I+FEER+
Sbjct: 499 KKPSVTIEFEERW 511
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 439

A DNA sequence (GBSx0475) was identified in *S.agalactiae* <SEQ ID 1405> which encodes the amino acid sequence <SEQ ID 1406>. This protein is predicted to be PblA. Analysis of this protein sequence reveals the following:

```
Possible site: 57
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -7.11   Transmembrane  427 - 443 ( 424 - 445)
   INTEGRAL    Likelihood = -4.99   Transmembrane  449 - 465 ( 448 - 469)
   INTEGRAL    Likelihood = -2.71   Transmembrane   41 -  57 (  38 -  57)
   INTEGRAL    Likelihood = -0.37   Transmembrane  361 - 377 ( 361 - 377)
   INTEGRAL    Likelihood = -0.22   Transmembrane  324 - 340 ( 324 - 340)

----- Final Results -----
               bacterial membrane --- Certainty=0.3845(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAG18638 GB:AY007505 PblA [Streptococcus mitis]
 Identities = 233/401 (58%), Positives = 296/401 (73%), Gaps = 17/401 (4%)

Query: 1   MATNLGQAYVQIMPSAKGISGSISKTLDPEASSAGSSAGSLLGGKLIGILGSVIAAAKIG 60
           MAT + QAYVQ++PSA+GI+G I   L+PEAS+AG SAG  LG  L+G++  VIAAA IG
```

```
Sbjct: 1   MATEIAQAYVQLIPSARGITGKIQSILNPEASAAGQSAGQSLGSSLVGVMTKVIAAAGIG 60

Query: 61  EMVTKAISSSISEGAALQQSLGGVETLFKSNANLVKKYADEAYKTTGLSANAYMESVTGF 120
               KA S++ISEGAALQQSLGG+ETLFK +A+ VK YA+EAYKTTGLSANAYME+VTGF
Sbjct: 61  ----KAFSAAISEGAALQQSLGGIETLFKGSADKVKGYANEAYKTTGLSANAYMENVTGF 116

Query: 121 SASLLQSLGGDTAKAAKVANMAMIDMADNSNKMGTSMESIQYAYQGFAKQNYTMLDNLKL 180
           SASLLQSLGGDT KAA+ ANMAMIDM+DN+NKMGTSMESIQ AYQGFAKQNYTMLDNLKL
Sbjct: 117 SASLLQSLGGDTNKAAETANMAMIDMSDNANKMGTSMESIQMAYQGFAKQNYTMLDNLKL 176

Query: 181 GYGGTQEEMKRLLSDAQKLTGKKYDISNLSDVYEAIHAIQGKIGITGTTAKEAATTFTGS 240
           GYGGT++EM+RLL+DA+KLTG KYDI+NLSDVY AIHAIQ  + ITGTTAKEAA+TF+GS
Sbjct: 177 GYGGTKQEMQRLLADAEKLTGVKYDINNLSDVYSAIHAIQENLDITGTTAKEAASTFSGS 236

Query: 241 FEAMKAASKNLLGKMALGEDIKPSLKALFDTTSNFVLNNFIPMLTNVFKGFGSVISLTFS 300
           FE+MKAA++N+LGK+ALGE+I PSL AL  TTS F+ +NF+PM+ NVF G G V++   S
Sbjct: 237 FESMKAAAQNVLGKLALGENILPSLHALLKTTSTFLFDNFLPMIGNVFSGLGLVLTEGIS 296

Query: 301 ELIPKIVGFMQTSGPSLMQSGISFIISFVNGFLTAYPAFLTVAGKIFTDFVSFVMQSIPG 360
           ++  ++ G           S +   +S + G    +  F  + G +       ++ +I G
Sbjct: 297 QIASQLFG-------DAFGSAVFDQLSRITGIFETF--FDMIFGSLSKQDNIDILNTI-G 346

Query: 361 LLQAGATLVLNLIDGILANLPQIATS---AVSVISSFISML 398
             +  AT ++N+ D I      I ++    V ++  F+  L
Sbjct: 347 FSEEAATQIVNIADNIRVTFENIGSAIGDVVGIVGDFVGDL 387
 Identities = 112/386 (29%), Positives = 172/386 (44%), Gaps = 18/386 (4%)

Query: 235 TTFTGSFEAMKAASKNLLGKMA-LGEDIKPSLKA---LFDTTSNFVLNNFIPMLTNVFKG 290
           TT+    E++KA    ++   + L E IK +      L  T    V+  FI    N++
Sbjct: 580 TTWNAYVESLKAMWNAVVTFFSDLWESIKEAASTAWTLITTAVMMVVQPFIDGFMNIWNN 639

Query: 291 FGSVISLTFSELIPKIVGFMQTSGPSLMQSGISFIISFVNGFLTAYPAFLTVAGKIFTDF 350
               ++  +  +     G  +    S+    I  II  V G      A L++  +   +
Sbjct: 640 ISEGLTQVWEGIKLIFEGAWEFI-KSIFLGAILIIIDLVTGNFGQLGADLSLIWEGIKNG 698

Query: 351 VSFVMQSIPGLLQAGATLVLNLIDGILANLPQIATSAVSVISSFISMLQANYPAILKKGF 410
           +S + + I         +++    G+  N   + ++    I +  SM    +  I
Sbjct: 699 ISLIWEGIKTYFSGVVDVIVGYATGVFENFSNVLSTIWEFIKTAASMA---WEWIKSTVS 755

Query: 411 EILSYLVQGIIARLPDIVITVGKL---IAILAGAIASNLPKVLALGVQLLITFVKGILSV 467
            +++ L+QG      + V  + L   I   A A  S L K L LG   +   V G  +
Sbjct: 756 NLITGLIQGAQNLWNNFVSFLSGLWENIKSTASAAWSGL-KSLVLG--FINGLVSGAQTA 812

Query: 468 IGKINETANNIGEK---LINAIKSIDLLSAGRAIMRGFLRGLEDVWGDIQNFVGDIAGWI 524
              + +  +++  K   + N IK+I+L  AG+AI+ GFL GL+  W  + NFVG IA WI
Sbjct: 813 WNNMKQAVSDLVTKVTNIFNGIKNINLWEAGKAILNGFLGGLKSAWEGVTNFVGGIANWI 872

Query: 525 KDHKGPISYDRRLLIPAGNAIMQGLHQGLVDKFKPVKNLVNGMAEEIQSSFGNPQLAFDM 584
           +DHKGPI YDR+LLIPAGNAIM  L  GL D FK VK  V GM+ EI   F    L  +
Sbjct: 873 RDHKGPIEYDRKLLIPAGNAIMGSLDNGLKDGFKDVKKTVGGMSGEISDVFSGDNLDLNS 932

Query: 585 DTNVNNGFE-RIGTLNKNLSSQVTST 609
             +V    E R+   +  L  Q + T
Sbjct: 933 TASVTKNLEARLAMPSAQLEVQESKT 958
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1407> which encodes the amino acid sequence <SEQ ID 1408>. Analysis of this protein sequence reveals the following:

```
     Possible site: 55
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -2.76    Transmembrane  458 - 474 ( 458 - 474)
   INTEGRAL    Likelihood = -2.60    Transmembrane  483 - 499 ( 482 - 499)
   INTEGRAL    Likelihood = -2.02    Transmembrane  429 - 445 ( 429 - 445)
   INTEGRAL    Likelihood = -1.28    Transmembrane  397 - 413 ( 397 - 413)
   INTEGRAL    Likelihood = -0.53    Transmembrane  739 - 755 ( 738 - 755)
   INTEGRAL    Likelihood = -0.27    Transmembrane  356 - 372 ( 356 - 372)
```

```
----- Final Results -----
              bacterial membrane --- Certainty=0.2105(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAB18717 GB:U38906 ORF42 [Bacteriophage r1t]
 Identities = 261/579 (45%), Positives = 359/579 (61%), Gaps = 63/579 (10%)

Query: 184 MKRLLSDAEKLPAAMGKKFDLSNYADVVEAIHLVQDNMGIAGVAAEEAKTTFSGSLAAMK 243
            M+RLL+DA+KL    G+K+D+SN++D+ +AIH +Q  M I G  A+EA TTFSGS  +MK
Sbjct: 1   MQRLLTDAQKLT---GQKYDISNFSDITQAIHAIQTEMDITGTTAKEASTTFSGSFDSMK 57

Query: 244 SSFTNVMAGLSLGDDIRPALRGLAETTSNFLFGNFIPMVANIFKGLPSAIGTFIGAAAPI 303
            ++ +NV+  LSLG D++  L  L  TTS FLF NFIPMV NIFK LP AI TF+ AA
Sbjct: 58  AAMSNVLGNLSLGRDLQGPLNALVSTTSTFLFKNFIPMVGNIFKALPGAISTFVSAAGKE 117

Query: 304 ITSQ-------------------------------------FQGLMSSLG-ISIDLSPIT 325
            ++SQ                                     F  L+SS+G IS  +  +
Sbjct: 118 LSSQLGNGIGSGFSDFTAKFSSILSPLQGSFQTIVSGLKPVFDSLLSSIGPISTQIMGVF 177

Query: 326 AKFAQIGQNLQ----PVFNGLKTAFSQLPSFFTSIGSAVAPVIDTIISGLARLDFSGFEA 381
            +K  Q+  N+     PV + L  AF QLPS F +I  AV P+IDTI SG++RLDFSG +A
Sbjct: 178 SKLPQLFSNVISAVIPVISTLSVAFGQLPSLFEAISVAVQPMIDTISSGISRLDFSGIQA 237

Query: 382 LISAILPALQAGFSNFAAIVGPAISGVVDSFVGMWNAAQPLISILSDALMPVFQILGSFL 441
            +ISA++PA+  G +    I+GP+I  +V+SFV MWN+ QPL ++++ ALMP FQ+LG+F+
Sbjct: 238 IISALVPAITTGITTMMGIIGPSIDTLVNSFVKMWNSIQPLATVIAGALMPAFQVLGAFI 297

Query: 442 GGVVKGALMGVSFAFDAVKVAIQLVTPIIDLLVQGLNFVQPVLSVIAEWIGVAIGMFGNL 501
            GGV+KGA++ +S  FD ++V +  +TPII  ++       PVL+ +A+W+G AIG F N
Sbjct: 298 GGVLKGAMLALSATFDTIRVVVGFLTPIIAAVLAKFQEFAPVLATVAQWVGTAIGFFANF 357

Query: 502 GTAGQGLSAFIKSAWTNIQTAISTAGTIISTVIDYIKLAFSGAGSAVGVLKNIFSLAWMA 561
            G AG  L   I SAW  I++ IS+  + I  +I+  K  F+G GSA G L+++ S AW
Sbjct: 358 GAAGTSLKGLITSAWNGIKSIISSVVSGIGGIINTAKAIFTGLGSAGGALRSMISGAWSG 417

Query: 562 MGDAINVAKGIISSVINGIKSAFSSFS-------SLVSSVGSAVNGVIDSISSTIRG--- 611
            +   I+   G IS  INGIKS FSS         S++S V S + G+I   SSTI G
Sbjct: 418 IRSIISSVGGSISGTINGIKSFFSSLGGSGNGLRSVMSGVWSGITGIISGASSTISGIID 477

Query: 612 --------LANIDISGAGAAIMNGFLNGLKSAWGAVKSFVSGIANWIAEHKGPISYDRVL 663
                    L NID++GAG A+++GF+ GLKS W A K FV GIA+WI +HKGPISYDR +
Sbjct: 478 GIKNIFNSLKNIDLAGAGRAVIDGFVGGLKSTWEAGKKFVGGIADWIKDHKGPISYDRKI 537

Query: 664 LKPAGKAIMGGLNTSLIDGFKEVKSNVSGMADDLASTMT 702
            L PAG+AIMGG N SL++ FK V+ NVSG+A  + S +T
Sbjct: 538 LIPAGQAIMGGFNDSLMENFKAVQKNVSGIAKQIQSAIT 576
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 272/701 (38%), Positives = 371/701 (52%), Gaps = 91/701 (12%)

Query: 1   MATNLGQAYVQIMPSAKGISGSISKTLDPEASSAGSSAGSLLGGKLIGILGSVIAAAKIG 60
            MAT LGQAYVQIMPSA+GISG+ISK LDPEA SAG SAGSL+GG L+ ++G  IAAA IG
Sbjct: 1   MATELGQAYVQIMPSARGISGAISKQLDPEARSAGLSAGSLIGGNLVKMIGGAIAAAGIG 60

Query: 61  EMVTKAISSSISEGAALQQSLGGVETLFKSNANLVKKYADEAYKTTGLSANAYMESVTGF 120
            +M    ISS++S GA LQQS GG++TL+K     VK +A EAYK  G+SAN Y E
Sbjct: 61  KM----ISSALSAGADLQQSFGGIDTLYKGAETAVKGFAKEAYKA-GISANTYAEQAVSM 115

Query: 121 SASLLQSLGGDTAKAAKVANMAMIDMADNSNKMGTSMESIQYAYQGFAKQNYTMLDNLKL 180
             ASL QSLGGD   AAK ANMA++DMADNS KMGT + SIQ AYQGFAKQNYTMLDNL+L
Sbjct: 116 GASLKQSLGGDAVAAAKAANMAIMDMADNSAKMGTDITSIQMAYQGFAKQNYTMLDNLRL 175

Query: 181 GYGGTQEEMKRLLSDAQKL---TGKKYDISNLSDVYEAIHAIQGKIGITGTTAKEAATTF 237
            GYGGT+EEMKRLLSDA+KL    GKK+D+SN +DV EAIH +Q  +GI G  A+EA TTF
```

```
Sbjct: 176 GYGGTKEEMKRLLSDAEKLPAAMGKKFDLSNYADVVEAIHLVQDNMGIAGVAAEEAKTTF 235

Query: 238 TGSFEAMKAASKNLLGKMALGEDIKPSLKALFDTTSNFVLNNFIPMLTNVFKGFGSVISL 297
           +GS  AMK++  N++  ++LG+DI+P+L+ L +TTSNF+  NFIPM+ N+FKG  S I
Sbjct: 236 SGSLAAMKSSFTNVMAGLSLGDDIRPALRGLAETTSNFLFGNFIPMVANIFKGLPSAIGT 295

Query: 298 TFSELIPKIV----GFMQTSGPSLMQSGISFIISFV--------NGFLTAY---PAFLTV 342
                 P I     G M + G S+  S I+   + +        NG  TA+   P+F T
Sbjct: 296 FIGAAAPIITSQFQGLMSSLGISIDLSPITAKFAQIGQNLQPVFNGLKTAFSQLPSFFTS 355

Query: 343 AGKIFTDFVSFVMQSIPGL----LQAGATLVLNLIDGILANLPQIATSAVS-VISSFISM 397
            G      +  ++  +  L     +A  + +L  +    +N   I   A+S V+ SF+ M
Sbjct: 356 IGSAVAPVIDTIISGLARLDFSGFEALISAILPALQAGFSNFAAIVGPAISGVVDSFVGM 415

Query: 398 LQANYPAI------LKKGFEILSYLVQGI---------------------IARLPDIVIT 430
             A  P I      L   F+IL   + G+                     +  + D+++
Sbjct: 416 WNAAQPLISILSDALMPVFQILGSFLGGVVKGALMGVSFAFDAVKVAIQLVTPIIDLLVQ 475

Query: 431 ----VGKLIAILAGAIASNLPKVLALGV--QLLITFVKGILSVIGKINETANNIGEKLIN 484
               V  +++++A  I   +     LG   Q L  F+K   + I     TA  I   +I+
Sbjct: 476 GLNFVQPVLSVIAEWIGVAIGMFGNLGTAGQGLSAFIKSAWTNIQTAISTAGTIISTVID 535

Query: 485 AIKSI-----------------------DLLSAGRAIMRGFLRGLEDVWGDIQNFVGDIA 521
            IK                         D ++  + I+   + G++  +    + V  +
Sbjct: 536 YIKLAFSGAGSAVGVLKNIFSLAWMAMGDAINVAKGIISSVINGIKSAFSSFSSLVSSVG 595

Query: 522 GWIKDHKGPISYDRRLLI-----PAGNAIMQGLHQGLVDKFKPVKNLVNGMAEEIQSSFG 576
             +      IS   R L       AG AIM G   GL   +  VK+ V+G+A  I    G
Sbjct: 596 SAVNGVIDSISSTIRGLANIDISGAGAAIMNGFLNGLKSAWGAVKSFVSGIANWIAEHKG 655

Query: 577 NPQLAFDMDTNVNNGFERIGTLNKNLSSQVTSTDNYTSGNA 617
              +++D       G   +G LN +L        +  SG A
Sbjct: 656 --PISYDRVLLKPAGKAIMGGLNTSLIDGFKEVKSNVSGMA 694
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 440

A DNA sequence (GBSx0477) was identified in *S.agalactiae* <SEQ ID 1409> which encodes the amino acid sequence <SEQ ID 1410>. Analysis of this protein sequence reveals the following:

```
Possible site: 16
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2565(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAG18637 GB:AY007505 unknown [Streptococcus mitis]
 Identities = 64/119 (53%), Positives = 87/119 (72%), Gaps = 2/119 (1%)

Query: 1   MLKMDEDALVCDLAETYHIYDYKQLPPLKVAVFSLGLREESRINRVISGNRVSFERRILA 60
           M++ DEDAL+CDLAETY I+DY+QLP  +VAVF+ GLR++SRI   ++ ++V FE  +LA
Sbjct: 1   MIQTDEDALICDLAETYGIFDYRQLPADQVAVFAFGLRDDSRIKLAMTNSKVPFETFLLA 60

Query: 61  GMFDRLGMLIWMKTTDGQKGKNRPEMVSTMF--DNQQKDSEVVSFGSGKDFEETRNNIL 117
           G+ DRL  L+W KTTDGQKG N+P MV+       + K+S+ + F SG+DFEE R  IL
Sbjct: 61  GVLDRLSALVWFKTTDGQKGINKPLMVTEELTGKTKAKESKEMIFDSGEDFEEYRQKIL 119
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1411> which encodes the amino acid sequence <SEQ ID 1412>. Analysis of this protein sequence reveals the following:

```
Possible site: 41
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2905(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 60/123 (48%), Positives = 82/123 (65%), Gaps = 2/123 (1%)

Query: 1    MLKMDEDALVCDLAETYHIYDYKQLPPLKVAVFSLGLREESRINRVISGNRVSFERRILA 60
            M+  D+DAL CDLAETY IYDY+QLP  +VAVF++GLR  SRI   +SG   + +  +LA
Sbjct: 1    MIAKDDDALTCDLAETYGIYDYRQLPAYQVAVFAVGLRSNSRIKMALSGETEALDTVLLA 60

Query: 61   GMFDRLGMLIWMKTTDGQKGKNRPEMV--STMFDNQQKDSEVVSFGSGKDFEETRNNILG 118
            G++D   +L W KT DGQ G+N+P+ V  +      QK ++V+SF SG+DFE  R  +LG
Sbjct: 61   GIYDNTNLLFWSKTKDGQSGQNKPKSVVEAISGSKSQKANDVISFVSGEDFENARKQLLG 120

Query: 119 FGG 121
              G
Sbjct: 121 GDG 123
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 441

A DNA sequence (GBSx0478) was identified in *S.agalactiae* <SEQ ID 1413> which encodes the amino acid sequence <SEQ ID 1414>. Analysis of this protein sequence reveals the following:

```
Possible site: 36
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2280(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAG18636 GB:AY007505 unknown [Streptococcus mitis]
 Identities = 40/80 (50%), Positives = 62/80 (77%), Gaps = 1/80 (1%)

Query: 3    TSSGFEYKIEESRLKNYELVEALADLESNPLSLPKVLRLLLGDQVESLKNHLRASDGTVS 62
            TS+GF ++I + RL+NYEL+EA++++++NP  LPKV++L+LG++ E LKNH+R +DG V
Sbjct: 24   TSTGFPFEITKERLENYELLEAISEVDTNPAVLPKVVKLMLGNKSEDLKNHVRTADGIVP 83

Query: 63   TEALMEEVKEIFES-GQLKK 81
             + +  E+ EIF S  QLKK
Sbjct: 84   LDKMGAEISEIFSSQNQLKK 103
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1415> which encodes the amino acid sequence <SEQ ID 1416>. Analysis of this protein sequence reveals the following:

```
Possible site: 56
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4365(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 42/75 (56%), Positives = 60/75 (80%)

Query: 2  KTSSGFEYKIEESRLKNYELVEALADLESNPLSLPKVLRLLLGDQVESLKNHLRASDGTV 61
           KT+SGFEY+I + RLKN+ELVEA+A+ E++P ++ K++ LLLGD  +SLK H+R ++G V
Sbjct: 7  KTTSGFEYEIPKKRLKNFELVEAIAEEETDPTAVVKIVNLLLGDAAKSLKEHVRDAEGIV 66

Query: 62 STEALMEEVKEIFES 76
             EA+  E+KEIFES
Sbjct: 67 DVEAIGVEIKEIFES 81
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 442

A DNA sequence (GBSx0479) was identified in *S.agalactiae* <SEQ ID 1417> which encodes the amino acid sequence <SEQ ID 1418>. This protein is predicted to be Structural protein. Analysis of this protein sequence reveals the following:

```
Possible site: 44
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.3461(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAG18635 GB:AY007505 unknown [Streptococcus mitis]
 Identities = 114/183 (62%), Positives = 142/183 (77%)

Query: 2    VANSSNVTTAKPKIGGAIYTAPLGTELPKDTASELNEAFKSLGYISEDGLSNEDKRESEE 61
             +A  +NVTTAKPKIGGA+Y+APLGT LP D  ++L++AF++LGYIS+DG++N +  ESE
Sbjct: 1    MATEANVTTAKPKIGGAVYSAPLGTALPTDATTKLDQAFEALGYISDDGMTNSNSPESEN 60

Query: 62   IQAWGGDVVESAQKSKADKFTYTLIEALNIEVLKEIYGKDNVTGDLKTGITVKSNSKPLE 121
             I+AWGG VV S QK K D F Y LIEALN+ VLKE+YG DNV+GDL +GIT+K+NSK L
Sbjct: 61   IKAWGGVVVSSVQKEKTDTFKYMLIEALNLHVLKEVYGPDNVSGDLSSGITIKANSKELP 120

Query: 122 EHCLVIEMILKNNTVKRIVIPKGKVSEVGEIKYVDNEAAGYETTLQAFPDAEGNTHYEYI 181
              HCLVIE +LK   +KRIVIP GKV+ + EI Y D    GY TT+ AFP+A  +THYEYI
Sbjct: 121 HHCLVIETVLKGGVLKRIVIPSGKVTAIDEITYNDGSVLGYGTTVTAFPNAADDTHYEYI 180

Query: 182 KGA 184
            KGA
Sbjct: 181 KGA 183
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1419> which encodes the amino acid sequence <SEQ ID 1420>. Analysis of this protein sequence reveals the following:

```
Possible site: 13
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.2379(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 119/182 (65%), Positives = 142/182 (77%)
```

```
Query: 4   NSSNVTTAKPKIGGAIYTAPLGTELPKDTASELNEAFKSLGYISEDGLSNEDKRESEEIQ 63
           ++ NVT+AKPK GGAIY+APLGTELPKD  SELN  FK+LGY+SEDG+ NED R SE I+
Sbjct: 6   DTKNVTSAKPKTGGAIYSAPLGTELPKDAKSELNTKFKNLGYVSEDGVVNEDTRSSENIK 65

Query: 64  AWGGDVVESAQKSKADKFTYTLIEALNIEVLKEIYGKDNVTGDLKTGITVKSNSKPLEEH 123
           AWGGD+V + Q  K DKFTY LIE+LN+EVLKE+YG  NVTGDL  GI +KSNSK LE H
Sbjct: 66  AWGGDIVGAVQTEKEDKFTYKLIESLNVEVLKEVYGAVNVTGDLSGGIQIKSNSKELEAH 125

Query: 124 CLVIEMILKNNTVKRIVIPKGKVSEVGEIKYVDNEAAGYETTLQAFPDAEGNTHYEYIKG 183
            +V++MI+    +KRIV+P  KV EVGEIKYVD E  GYETTL+ FPD +G+TH EYI
Sbjct: 126 VIVVDMIMNGGILKRIVLPNAKVDEVGEIKYVDGEVVGYETTLKCFPDKDGDTHREYIVK 185

Query: 184 AG 185
            G
Sbjct: 186 PG 187
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 443

A DNA sequence (GBSx0480) was identified in *S.agalactiae* <SEQ ID 1421> which encodes the amino acid sequence <SEQ ID 1422>. Analysis of this protein sequence reveals the following:

```
Possible site: 58
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2214(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAB18710 GB:U38906 ORF35 [Bacteriophage r1t]
 Identities = 52/78 (66%), Positives = 66/78 (83%)

Query: 1   MSKFKFKLNKAGVAELMKSSEMQQVLTTKATAIRERCGDGYAQDIHVGKNRANAMVSAKT 60
           M+K  FKLN++GVA +MKS EMQ +L  KA+A+++RCG GY QD+HVGKNRANAMV A+T
Sbjct: 1   MAKNLFKLNRSGVASMMKSPEMQAILKEKASAVKQRCGPGYGQDMHVGKNRANAMVFAET 60

Query: 61  IKAKKDNSKNNTLLKAVR 78
            +AK+DN KNNT+LKAVR
Sbjct: 61  YQAKRDNMKNNTILKAVR 78
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1423> which encodes the amino acid sequence <SEQ ID 1424>. Analysis of this protein sequence reveals the following:

```
Possible site: 54
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2446(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 75/78 (96%), Positives = 76/78 (97%)

Query: 1   MSKFKFKLNKAGVAELMKSSEMQQVLTTKATAIRERCGDGYAQDIHVGKNRANAMVSAKT 60
           MSKFKFKLN+AGVAELMKSSEMQQVLTTKATAIRERCGDGY QDIHVGKNRANAMVS KT
Sbjct: 1   MSKFKFKLNRAGVAELMKSSEMQQVLTTKATAIRERCGDGYVQDIHVGKNRANAMVSTKT 60
```

```
Query: 61 IKAKKDNSKNNTLLKAVR 78
          IKAKKDNSKNNTLLKAVR
Sbjct: 61 IKAKKDNSKNNTLLKAVR 78
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 444

A DNA sequence (GBSx0481) was identified in *S.agalactiae* <SEQ ID 1425> which encodes the amino acid sequence <SEQ ID 1426>. Analysis of this protein sequence reveals the following:

```
Possible site: 38
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2888(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAB18709 GB:U38906 ORF34 [Bacteriophage r1t]
 Identities = 41/59 (69%), Positives = 45/59 (75%)

Query: 1   MTGKKVEYILAIPKGDKHDWEDKEVCFFDKKWRTVGLALEGIEELIPLEWNKKVMVERY 59
            +TGKK  Y LAIPK D HDWE+K+V FF K WRT G  LEGIE LIPL+WNKKV VE Y
Sbjct: 56  LTGKKAIYTLAIPKKDTHDWENKKVRFFGKTWRTFGEPLEGIEGLIPLDWNKKVTVEHY 114
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1427> which encodes the amino acid sequence <SEQ ID 1428>. Analysis of this protein sequence reveals the following:

```
Possible site: 39
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2779(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 51/60 (85%), Positives = 57/60 (95%)

Query: 1   MTGKKVEYILAIPKGDKHDWEDKEVCFFDKKWRTVGLALEGIEELIPLEWNKKVMVERYE 60
            +TGKKVEY+LAIPKGD+HDWE+KEV FF KKWRTVG+ LEGIEELIPL+WNKKVMVERYE
Sbjct: 50  LTGKKVEYVLAIPKGDEHDWENKEVRFFGKKWRTVGIPLEGIEELIPLDWNKKVMVERYE 109
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 445

A DNA sequence (GBSx0482) was identified in *S.agalactiae* <SEQ ID 1429> which encodes the amino acid sequence <SEQ ID 1430>. Analysis of this protein sequence reveals the following:

```
Possible site: 25
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2770(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAB18708 GB:U38906 ORF33 [Bacteriophage r1t]
 Identities = 89/130 (68%), Positives = 106/130 (81%), Gaps = 1/130 (0%)

Query: 1   MTNFATTDDVILLWRQLSVDEIKRAEALLETVSDTLRLEASKVGKNLDEMILETP-YFAT 59
            M  FAT DD+ +LWR L  DE +RAE LLE VSD+LR EA KVG++L  MI E P YFA+
Sbjct: 1   MNPFATVDDLTMLWRPLKGDEKERAEKLLEIVSDSLREEADKVGRDLYAMIAEKPSYFAS 60

Query: 60  VLKSVTVDIVARTLMTATQGEPMSQESQSALGYTWSGTYLVPGGGLFIKDSELKRLGLKK 119
            V+KSVTVDIVARTLMT+T  EPM+Q ++SALGY+ SG+YLVPGGGLFIK+SEL RLGLKK
Sbjct: 61  VVKSVTVDIVARTLMTSTDQEPMTQTTESALGYSVSGSYLVPGGGLFIKNSELSRLGLKK 120

Query: 120 QRYGGIELYG 129
            QR+G I+ YG
Sbjct: 121 QRFGVIDFYG 130
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1431> which encodes the amino acid sequence <SEQ ID 1432>. Analysis of this protein sequence reveals the following:

```
Possible site: 37
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2061(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 116/138 (84%), Positives = 129/138 (93%)

Query: 3   NFATTDDVILLWRQLSVDEIKRAEALLETVSDTLRLEASKVGKNLDEMILETPYFATVLK 62
            NFATTDDVILLWR LSVDE+KRA ALL+ VSDTLR+EA KVGK+LD+ +++ PYF  V+K
Sbjct: 3   NFATTDDVILLWRPLSVDELKRANALLKVVSDTLRMEADKVGKDLDKTMVDKPYFVNVIK 62

Query: 63  SVTVDIVARTLMTATQGEPMSQESQSALGYTWSGTYLVPGGGLFIKDSELKRLGLKKQRY 122
            SVTVDIVARTLMT+T+GEPM+QESQSALGYTWSGTYLVPGGGLFIKDSELKRLGLKKQRY
Sbjct: 63  SVTVDIVARTLMTSTRGEPMAQESQSALGYTWSGTYLVPGGGLFIKDSELKRLGLKKQRY 122

Query: 123 GGIELYGEIERNNSYFSR 140
            GGIELYGEIER+NS FSR
Sbjct: 123 GGIELYGEIERDNSCFSR 140
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 446

A DNA sequence (GBSx0483) was identified in *S.agalactiae* <SEQ ID 1433> which encodes the amino acid sequence <SEQ ID 1434>. This protein is predicted to be Structural protein. Analysis of this protein sequence reveals the following:

```
Possible site: 30
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3015(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAB18706 GB:U38906 Structural protein [Bacteriophage r1t]
 Identities = 132/296 (44%), Positives = 189/296 (63%), Gaps = 8/296 (2%)

Query: 5   IKAGTLFKPELVTEIMSKVKGHSTLAKLSGQTPIPFNGVEQFVFNLDGNAQIVGEGEQKL 64
            +  GTLF P LVT+++SKV G S++A+LS Q PIPFNG + F F +D    +V E  +K
Sbjct: 3   LNKGTLFDPTLVTDLISKVAGKSSIARLSAQKPIPFNGEKVFTFTMDSEIDVVAESGKKT 62

Query: 65  GNTAKVTSKIIKPLKFVYQARMTDEFKYASEEKRLNFLKHYADGFAKKMAEAFDIAAIHG 124
                 +  + + P+K  Y AR++DEF YAS+E+++N L+ + DGFAKK+A   D+ A HG
Sbjct: 63  HGGVTLAPQTMVPIKVEYGARISDEFMYASDEEKINILQEFNDGFAKKVARGIDLMAFHG 122

Query: 125 LEPRTMTDASFKATNSFDGVVTGNVIKYEADK--IDDN--IDAAVTTIVANGNDVTGIAL 180
            + PR  T ++   TN FD  VT    K EA +   D N  I+ AV  +     DVTGIA+
Sbjct: 123 VNPRLGTASAVIGTNHFDSKVTQ---KVEAPRGIADPNGAIENAVELLTGVDADVTGIAI 179

Query: 181 SPQAGQDMSKRKDKFDNVMYPEFRFGQRPSNFFNMTLDINKTLTMKGGTAKDDHAIVGDF 240
            +P     ++K+KD  DN ++PE ++G  P     + +D+NKT++    T + D AI+GDF
Sbjct: 180 NPSFRSALAKQKDLQDNALFPELKWGATPDTINGLPVDVNKTVSDMSLTQR-DRAIIGDF 238

Query: 241 QNMFKWGYAENIPMEIIEYGDPDGSGRDLKAYNEILLRTEAFIGWGILDEKAFSRV 296
             N FKWGYA+ +P+E+I+YGDPD SG DLK YN++ +R E F+GWGILD   F+RV
Sbjct: 239 ANGFKWGYAKEVPLEVIQYGDPDNSGLDLKGYNQVYIRAELFLGWGILDATKFARV 294
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1435> which encodes the amino acid sequence <SEQ ID 1436>. Analysis of this protein sequence reveals the following:

```
Possible site: 30
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2772(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 133/298 (44%), Positives = 187/298 (62%), Gaps = 2/298 (0%)

Query: 1   MAESIKAGTLFKPELVTEIMSKVKGHSTLAKLSGQTPIPFNGVEQFVFNLDGNAQIVGEG 60
            M       +LF   LV+++++KVKGHS+LAKLS Q PIPFNG ++F F LD +  +V E
Sbjct: 1   MGTETSKASLFDKHLVSDLINKVKGHSSLAKLSSQKPIPFNGSKEFTFTLDSDIDVVAEN 60

Query: 61  EQKLGNTAKVTSKIIKPLKFVYQARMTDEFKYASEEKRLNFLKHYADGFAKKMAEAFDIA 120
             +K      +    I P+K  Y AR++DEF YA+EE++++ LK + +GFAKK+A   D+
Sbjct: 61  GKKTHGGLSLEPVTIVPIKVEYGARLSDEFLYATEEEKIDILKAFNEGFAKKLARGIDLM 120

Query: 121 AIHGLEPRTMTDASFKATNSFDGVVTGNVIKYEADKIDDNIDAAVTTIVANGNDVTGIAL 180
            A+HG+ PRT   +    TN FD  VT  V   E++  D NI+AAV  I  +   VTG+A+
Sbjct: 121 AMHGINPRTKKASDVIGTNHFDSKVTQVVKFTESEDADANIEAAVNLIQGSEGVVTGLAM 180

Query: 181 SPQAGQDMSK-RKDKFDNVMYPEFRFGQRPSNFFNMTLDINKTLTMKGGTAKD-DHAIVG 238
              +    ++K     +   MYPE  +G  P +   +   +N T+      A+  D  I+G
Sbjct: 181 DTEFSTALAKVTNGEMGPKMYPELAWGANPDSINGLKSSVNTTVGAGADEAESKDLVIIG 240

Query: 239 DFQNMFKWGYAENIPMEIIEYGDPDGSGRDLKAYNEILLRTEAFIGWGILDEKAFSRV 296
            DF++MFKWGYA+ IPMEII+YGDPD SG+DLK YN+I LR EA+IGWGILD K+F+RV
Sbjct: 241 DFESMFKWGYAKQIPMEIIKYGDPDNSGKDLKGYNQIYLRAEAYIGWGILDAKSFARV 298
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 447

A DNA sequence (GBSx0484) was identified in *S.agalactiae* <SEQ ID 1437> which encodes the amino acid sequence <SEQ ID 1438>. Analysis of this protein sequence reveals the following:

```
Possible site: 61
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2224(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9659> which encodes amino acid sequence <SEQ ID 9660> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAB18705 GB:U38906 ORF30 [Bacteriophage r1t]
 Identities = 64/158 (40%), Positives = 101/158 (63%), Gaps = 8/158 (5%)

Query: 43  MSEFKVIETQEELDTIVKARIARERE----KYQDYDQLKTRVEELETENSSLQTALNDAK 98
            MSE  + +TQEEL+ I++ R+AR++E     + DYD+LKT++  LE +N++ Q  + ++K
Sbjct: 1   MSENNLPKTQEELNQIIETRLARQKETIEANFADYDELKTKIAALEADNTAYQATIEESK 60

Query: 99  SNTDSYTEKITTLENQIAGYEAANLRTKVALQYGLPIDLANRLQGDDEDGLKVDAERLAS 158
            S    + ++    E QI+GY+   L+  +A++ GLP+DLA+RL GDDE+ LK DAER +
Sbjct: 61  S----WEQEKADYEKQISGYKTTQLKQSIAIKAGLPLDLADRLSGDDEESLKADAERFSG 116

Query: 159 FIKPSQPQPPTKSNEPIITDQKEAGWIEMARNLVNKGE 196
            FIKP  P  P K  EP + D K+  + ++    L  +GE
Sbjct: 117 FIKPKTPPAPLKDVEPNLGDGKDGAYRKLVDGLKTEGE 154
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1439> which encodes the amino acid sequence <SEQ ID 1440>. Analysis of this protein sequence reveals the following:

```
Possible site: 59
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3476(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 128/149 (85%), Positives = 136/149 (90%)

Query: 43  MSEFKVIETQEELDTIVKARIAREREKYQDYDQLKTRVEELETENSSLQTALNDAKSNTD 102
            MSEFKVIETQEELDTIVKARIAREREKYQDYDQLKTRVEELETENSSLQTALNDAKSNTD
Sbjct: 1   MSEFKVIETQEELDTIVKARIAREREKYQDYDQLKTRVEELETENSSLQTALNDAKSNTD 60

Query: 103 SYTEKITTLENQIAGYEAANLRTKVALQYGLPIDLANRLQGDDEDGLKVDAERLASFIKP 162
            SYTE+I+TL+NQIA YE ANLRTKVALQYGLPIDLA+RLQGDDEDGLKVDAERLASFIKP
Sbjct: 61  SYTEEISTLKNQIADYETANLRTKVALQYGLPIDLADRLQGDDEDGLKVDAERLASFIKP 120

Query: 163 SQPQPPTKSNEPIITDQKEAGWIEMARNL 191
            SQPQPP KSNEP I    +A +  + + L
Sbjct: 121 SQPQPPAKSNEPNIDSNADANYRALVQGL 149
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 448

A DNA sequence (GBSx0485) was identified in *S.agalactiae* <SEQ ID 1441> which encodes the amino acid sequence <SEQ ID 1442>. Analysis of this protein sequence reveals the following:

```
Possible site: 56
```

```
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2888(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAB18704 GB:U38906 ORF29 [Bacteriophage r1t]
 Identities = 322/461 (69%), Positives = 383/461 (82%)

Query: 8   KLGNQRPTQSVNLHFAKTLAHEAINYYKKTGLSCYLWQENMLIPMMAINEDNLWVHQKYG 67
           + GNQ PTQSV L F +T   EAI  Y+K+   CY WQ+N+L  +MAI+ED LW HQK+G
Sbjct: 6   RFGNQYPTQSVILPFTETKYQEAIEIYEKSKHECYPWQKNLLKEVMAIDEDGLWTHQKFG 65

Query: 68  YAIPRRNGKTEVVYILELWALHKGLKILHTAHRISTSHSSFEKVKKYLEMSGYVDGEDFI 127
           Y+IPRRNGKTE+VYILELW+L +GL ILHTAHRISTSHSS+EK+KKYLE SGYV+GEDF
Sbjct: 66  YSIPRRNGKTEIVYILELWSLVQGLSILHTAHRISTSHSSYEKLKKYLEDSGYVEGEDFK 125

Query: 128 SNKAKGQERIEFKSSGSVIQFRTRTSNGGLGEGFDLLIIDEAQEYTAEQESALKYTVTDS 187
           S KAKGQER+E   SG VIQFRTRTS+GGLGEGFD+L+IDEAQEYT EQESALKYTVTDS
Sbjct: 126 SIKAKGQERLELIESGGVIQFRTRTSSGGLGEGFDILVIDEAQEYTTEQESALKYTVTDS 185

Query: 188 DNPMTIMCGTPPTMVSTGTVFESYRKECLKGDRRYSGWAEWSVDEMQPIHDVKSWYVANP 247
           DNPMTIMCGTPPT VS+GTVF +YR   + G  +YSGWAEWSV++++ IHDV++WY +NP
Sbjct: 186 DNPMTIMCGTPPTPVSSGTVFTNYRDNTIAGKAKYSGWAEWSVEDVKDIHDVEAWYNSNP 245

Query: 248 SMGYHLNERKIEAELGEDEIDHNIQRLGYWPSFNQKSVISEKEWAKLKVEQVPELKSKLF 307
           SMGYHLNERKIEAELGED++DHN+QRLGYWP +NQKSVISE+EW  LKV ++P +K KLF
Sbjct: 246 SMGYHLNERKIEAELGEDKLDHNVQRLGYWPKYNQKSVISEQEWNALKVNRLPVIKGKLF 305

Query: 308 VGIKFGQDGNNVSLSIAARASENKVFVEAIDCLSVRNGTQWIINFLKSADIAKVVVDGAS 367
           VGIK+G DG NV++SIA +    KVFVE IDC S+RNG QWIINFLK AD+ KVV+DG S
Sbjct: 306 VGIKYGNDGANVAMSIAVKTLSGKVFVETIDCQSIRNGNQWIINFLKKADVEKVVIDGQS 365

Query: 368 GQELLAQEMREHGLKKPELPKVAEIITANTMWEQGIMQETICHNDQPSLTAVVTNCEKRQ 427
           GQ +L  EM++  LK+P LP V EII AN++WEQGI Q+  CH+ QPSL+ VVTNC+KR
Sbjct: 366 GQSILTSEMKDFKLKEPILPTVKEIINANSLWEQGIFQKNFCHSGQPSLSTVVTNCDKRN 425

Query: 428 IGSNGGFGYKSLYDDRDISLMDSALLAHWICYTTKPKRKQR 468
           IG++GGFGYKS +DD DISLMDSALLAHW C   KPK+KQ+
Sbjct: 426 IGTSGGFGYKSQFDDMDISLMDSALLAHWACSNNKPKKKQQ 466
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1443> which encodes the amino acid sequence <SEQ ID 1444>. Analysis of this protein sequence reveals the following:

```
Possible site: 32
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3133(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 437/471 (92%), Positives = 459/471 (96%)

Query: 1   MVTKTKAKLGNQRPTQSVNLHFAKTLAHEAINYYKKTGLSCYLWQENMLIPMMAINEDNL 60
           MVTKTK KLGNQRPTQSVNLHFAK+LAHEAINYYKKTGLSCY WQ NMLIP+MAI+E+ L
Sbjct: 6   MVTKTKTKLGNQRPTQSVNLHFAKSLAHEAINYYKKTGLSCYPWQVNMLIPIMAIDENGL 65

Query: 61  WVHQKYGYAIPRRNGKTEVVYILELWALHKGLKILHTAHRISTSHSSFEKVKKYLEMSGY 120
           WVHQKYGYAIPRRNGKTEVVYI++LWALHKGLKILHTAHRISTSH+SFEKVKKYLEMSGY
Sbjct: 66  WVHQKYGYAIPRRNGKTEVVYIVQLWALHKGLKILHTAHRISTSHASFEKVKKYLEMSGY 125
```

```
Query: 121 VDGEDFISNKAKGQERIEFKSSGSVIQFRTRTSNGGLGEGFDLLIIDEAQEYTAEQESAL 180
           VDGEDFISNKAKGQERIEFK+SG+VIQFRTRTSNGGLGEGFDLLIIDEAQEYT+EQESAL
Sbjct: 126 VDGEDFISNKAKGQERIEFKASGAVIQFRTRTSNGGLGEGFDLLIIDEAQEYTSEQESAL 185

Query: 181 KYTVTDSDNPMTIMCGTPPTMVSTGTVFESYRKECLKGDRRYSGWAEWSVDEMQPIHDVK 240
           KYTVTDSDNPMTIMCGTPPTMVSTGTVFE+YRK+CLKG++RYSGWAEWSV EM  I+DV
Sbjct: 186 KYTVTDSDNPMTIMCGTPPTMVSTGTVFEAYRKDCLKGNKRYSGWAEWSVPEMVKINDVS 245

Query: 241 SWYVANPSMGYHLNERKIEAELGEDEIDHNIQRLGYWPSFNQKSVISEKEWAKLKVEQVP 300
           SWY++NPSMG+HLNERKIEAELGEDEIDHNIQRLGYWPSFNQKSVISEKEWAKLKVEQVP
Sbjct: 246 SWYISNPSMGFHLNERKIEAELGEDEIDHNIQRLGYWPSFNQKSVISEKEWAKLKVEQVP 305

Query: 301 ELKSKLFVGIKFGQDGNNVSLSIAARASENKVFVEAIDCLSVRNGTQWIINFLKSADIAK 360
           ELKSKLFVGIKFGQDGNNVSLSIAAR SENKVFVE IDCLSVRNGTQWIINFLKSADIAK
Sbjct: 306 ELKSKLFVGIKFGQDGNNVSLSIAARTSENKVFVETIDCLSVRNGTQWIINFLKSADIAK 365

Query: 361 VVVDGASGQELLAQEMREHGLKKPELPKVAEIITANTMWEQGIMQETICHNDQPSLTAVV 420
           VV+DGASGQELLAQEM++ GLKKPELPKVAEIITAN MWEQGIMQETICH+DQPSLTAVV
Sbjct: 366 VVIDGASGQELLAQEMKDQGLKKPELPKVAEIITANMMWEQGIMQETICHSDQPSLTAVV 425

Query: 421 TNCEKRQIGSNGGFGYKSLYDDRDISLMDSALLAHWICYTTKPKRKQRTSC 471
           TNCEKRQIGSNGGFGYKSLYDDRDISLMDSALLAHWICYTTKPKRKQRTSC
Sbjct: 426 TNCEKRQIGSNGGFGYKSLYDDRDISLMDSALLAHWICYTTKPKRKQRTSC 476
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 449

A DNA sequence (GBSx0486) was identified in *S.agalactiae* <SEQ ID 1445> which encodes the amino acid sequence <SEQ ID 1446>. Analysis of this protein sequence reveals the following:

```
Possible site: 32
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2745(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 450

A DNA sequence (GBSx0487) was identified in *S.agalactiae* <SEQ ID 1447> which encodes the amino acid sequence <SEQ ID 1448>. Analysis of this protein sequence reveals the following:

```
Possible site: 32
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2568(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAB18703 GB:U38906 ORF28 [Bacteriophage r1t]
```

```
 Identities = 124/250 (49%), Positives = 164/250 (65%), Gaps = 3/250 (1%)

Query: 2   VDDVLPKLLKSVQQDFEKHFGKSEVVAKAFAELQAKKATYKTVNEFAVEVGRLLSLALAN 61
           ++D+LP LL+ + QDF++    S+ + ++   L+ KKATY   NEF VEVG++LS  L
Sbjct: 1   MEDILPPLLEKINQDFDERAANSKKLKQSMELLKTKKATYIQANEFGVEVGQILSDVLGT 60

Query: 62  SVISDELPDGKMYYNIANRLVNDTLRHNYKLISDYAGDVQQNLNKQAKISLKIQRPPLNQ 121
            V  D LPDGKMY+NIA+RL+N  L+ N+ LIS Y+ DVQ  LN+ A   LK Q P LNQ
Sbjct: 61  HVTVDVLPDGKMYFNIADRLLNSILKKNFDLISGYSTDVQSELNQLAGFKLKSQVPELNQ 120

Query: 122 DKIDGLVNRLASEPVFDDVKWLLDEPIVNFSQSIVDDCIRANADFHFKTGLKPTIERIST 181
           D+IDG+VNR++SE  F+ + WLL EPIV FSQS+VDD ++ N DF  K GLKP I R
Sbjct: 121 DRIDGIVNRISSEDDFEKILWLLKEPIVTFSQSVVDDTLKKNIDFQAKAGLKPKIVRKLV 180

Query: 182 GKCCDWCDRLAGRYVYHEEPKDFYKRHQHCQCVIDYHPK--NGKRQNSWSKKWTKETTDI 239
           GK CDWC  LAG Y Y   P D Y RH+ C+C ++Y P+  + KRQ+ WSK W     D
Sbjct: 181 GKACDWCRNLAGSYDYPNVPSDVYHRHERCRCTVEYDPRDIDKKRQDVWSKNWVDPDKDA 240

Query: 240 -LERRKQMNI 248
            +  RK +N+
Sbjct: 241 KIAERKNLNL 250
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1449> which encodes the amino acid sequence <SEQ ID 1450>. Analysis of this protein sequence reveals the following:

```
Possible site: 32
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3099(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 169/261 (64%), Positives = 207/261 (78%), Gaps = 2/261 (0%)

Query: 1   MVDDVLPKLLKSVQQDFEKHFGKSEVVAKAFAELQAKKATYKTVNEFAVEVGRLLSLALA 60
           MVDDVLPKLLKSV+QDFEK+FG+S+VV KAFAELQAKK TYKTVNEFA+EVGRLLSLAL
Sbjct: 1   MVDDVLPKLLKSVRQDFEKYFGESDVVTKAFAELQAKKVTYKTVNEFAIEVGRLLSLALT 60

Query: 61  NSVISDELPDGKMYYNIANRLVNDTLRHNYKLISDYAGDVQQNLNKQAKISLKIQRPPLN 120
            SV SD+LPDGKMYYNIA RL+++T+  NYKLIS YAGDVQ+ LN+ A+I LK+QRPPLN
Sbjct: 61  GSVSSDKLPDGKMYYNIAKRLLDETMGRNYKLISGYAGDVQRILNENAQIGLKVQRPPLN 120

Query: 121 QDKIDGLVNRLASEPVFDDVKWLLDEPIVNFSQSIVDDCIRANADFHFKTGLKPTIERIS 180
           +DKI+G+VNRL SE  FDDVKWL  EPIVNFSQSIVDD I+ANAD  +KTG+ P + R
Sbjct: 121 RDKINGMVNRLDSENTFDDVKWLFGEPIVNFSQSIVDDTIKANADLQYKTGMTPQVVRTE 180

Query: 181 TGKCCDWCDRLAGRYVYHEEPKDFYKRHQHCQCVIDYHPKNGKRQNSWSKKWTK--ETTD 238
           +G CC+WC  + G Y Y + PKD ++RHQ C+C +DY PKNGK Q++WSK W K  +T +
Sbjct: 181 SGNCCEWCREVVGTYSYPKVPKDVWRRHQRCRCTLDYDPKNGKVQSAWSKIWRKKEKTQE 240

Query: 239 ILERRKQMNIDIRDNNRKSDI 259
            +ER ++        + K+DI
Sbjct: 241 SIERVEKFKESALVESIKNDI 261
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 451

A DNA sequence (GBSx0488) was identified in *S.agalactiae* <SEQ ID 1451> which encodes the amino acid sequence <SEQ ID 1452>. This protein is predicted to be Structural protein. Analysis of this protein sequence reveals the following:

```
Possible site: 58
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.38    Transmembrane   93 - 109 (   93 - 110)

----- Final Results -----
               bacterial membrane --- Certainty=0.1553(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAC39307 GB:AF022773 ORF5 [Lactococcus bacteriophage phi31]
 Identities = 271/410 (66%), Positives = 326/410 (79%), Gaps = 2/410 (0%)

Query: 1   MNYMGMGYLQRKLALFKTGVDKRYRYYAMDDRDNTRSIVMPDNVREMYRSVIEWTAKGVD 60
           M   G+GYL+ KL++ K   + RY  YAM   D  + I +P  + + YRS++ W AKGVD
Sbjct: 1   MTEKGIGYLRFKLSVHKRRAEMRYEQYAMKHVDRFKGITIPQALSQQYRSILGWCAKGVD 60

Query: 61  SLADRIIFREFANDDFNAWEIFKANNPDIFFDTAIQSALIASCCFVYIMPGKEDSLPKMQ 120
           SLADR+IFREF NDDF   EIF+ NNPDIFFD+A+ SALIASC F+YI  G+ D++ ++Q
Sbjct: 61  SLADRLIFREFENDDFTVNEIFEENNPDIFFDSAVLSALIASCSFIYISKGENDAV-RLQ 119

Query: 121 VIEASKATGILDPTTFLLTEGYAVLESDSNENPTLEAYFTGEKTWYYPKDEKP-YSIDNS 179
           VIEA  ATGI+DP T LLTEGYAVLE D N N  LEA+F  ++T YY +D +   SI N
Sbjct: 120 VIEAVNATGIIDPITGLLTEGYAVLERDENNNVVLEAHFLPDRTDYYYRDSRNNISIANP 179

Query: 180 TGHPLLVPVIHRPDAVRPFGRSRITKAGMYHQKAAKRTLERAEVTAEFYSFPQKYVLGMD 239
           TGHPLLVP+IHRPDAVRPFGRSRIT++GMY Q  AKRTLERA+VTAEFYSFPQKYV G+
Sbjct: 180 TGHPLLVPIIHRPDAVRPFGRSRITRSGMYWQSNAKRTLERADVTAEFYSFPQKYVTGLS 239

Query: 240 PDAEPMEKWRATVSTLLEISKDEDGDKPTVGQFTTASMAPFMDHLKMYASLFAGGSGLTL 299
            DAEPME W+ATVS++L+ +KDEDGDKPT+GQFT  SM+PF + L+  A+ FAG +GLTL
Sbjct: 240 DDAEPMETWKATVSSMLQFTKDEDGDKPTLGQFTQPSMSPFTEQLRTAAAGFAGETGLTL 299

Query: 300 DDLGFPSDNPSSVEAIKAAHENLRAAGRKAQRSFSSGFLNVAYIAVCLRDDFPYLRNQFM 359
           DDLGF SDNPSSVEAIKA+HENLR AGRKAQRS  +G LNVAY+A CLRDD PYLR QF
Sbjct: 300 DDLGFVSDNPSSVEAIKASHENLRLAGRKAQRSLGAGLLNVAYLAACLRDDVPYLREQFS 359

Query: 360 DTEIKWEPLFEADANMLTLVGDGAIKLNQAIPGFMDADVIRDLTGVKGSD 409
            T+ KWEPLFEADA+ML+L+GDGAIKLNQAIP F++ D IRDLTG+KG++
Sbjct: 360 KTKPKWEPLFEADASMLSLIGDGAIKLNQAIPEFINKDTIRDLTGIKGAE 409
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1453> which encodes the amino acid sequence <SEQ ID 1454>. Analysis of this protein sequence reveals the following:

```
Possible site: 58
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.38    Transmembrane   93 - 109 (   93 - 110)

----- Final Results -----
               bacterial membrane --- Certainty=0.1553(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 395/422 (93%), Positives = 407/422 (95%)

Query: 1   MNYMGMGYLQRKLALFKTGVDKRYRYYAMDDRDNTRSIVMPDNVREMYRSVIEWTAKGVD 60
           MNYMGMGYL+RKLALFKTGVDKRYRYYAMDDRD+TRSIVMP+NVREMYRSV+EWTAKGVD
```

```
Sbjct: 1    MNYMGMGYLRRKLALFKTGVDKRYRYYAMDDRDDTRSIVMPNNVREMYRSVLEWTAKGVD 60

Query: 61   SLADRIIFREFANDDFNAWEIFKANNPDIFFDTAIQSALIASCCFVYIMPGKEDSLPKMQ 120
            SLADRIIFREF NDDFNAWEIFKANNPDIFFDTAIQSALIASCCFVYIMPG ED LPKMQ
Sbjct: 61   SLADRIIFREFTNDDFNAWEIFKANNPDIFFDTAIQSALIASCCFVYIMPGAEDGLPKMQ 120

Query: 121 VIEASKATGILDPTTFLLTEGYAVLESDSNENPTLEAYFTGEKTWYYPKDEKPYSIDNST 180
           VIEASKATGILDPTTFLLTEGYA+LESDSN NPTLEAYFT +  WYYPK  KPY+I N T
Sbjct: 121 VIEASKATGILDPTTFLLTEGYAILESDSNGNPTLEAYFTDKDIWYYPKKGKPYNIKNPT 180

Query: 181 GHPLLVPVIHRPDAVRPFGRSRITKAGMYHQKAAKRTLERAEVTAEFYSFPQKYVLGMDP 240
           GHPLLVP+IHRPDAVRPFGRSRITKAGMYHQKAAKRTLERAEVTAEFYSFPQKYVLGMDP
Sbjct: 181 GHPLLVPIIHRPDAVRPFGRSRITKAGMYHQKAAKRTLERAEVTAEFYSFPQKYVLGMDP 240

Query: 241 DAEPMEKWRATVSTLLEISKDEDGDKPTVGQFTTASMAPFMDHLKMYASLFAGGSGLTLD 300
           DAEPMEKWRATVSTLLEISKDEDGDKPTVGQFTTASMAPFM+HLKMYASLFAGGSGLTLD
Sbjct: 241 DAEPMEKWRATVSTLLEISKDEDGDKPTVGQFTTASMAPFMEHLKMYASLFAGGSGLTLD 300

Query: 301 DLGFPSDNPSSVEAIKAAHENLRAAGRKAQRSFSSGFLNVAYIAVCLRDDFPYLRNQFMD 360
           DLGFPSDNPSSVE+IKAAHENLRAAGRKAQRSFSSGFLNVAYIAVCLRD+FPYLRNQFMD
Sbjct: 301 DLGFPSDNPSSVESIKAAHENLRAAGRKAQRSFSSGFLNVAYIAVCLRDEFPYLRNQFMD 360

Query: 361 TEIKWEPLFEADANMLTLVGDGAIKLNQAIPGFMDADVIRDLTGVKGSDNPIPKATEVTT 420
           T IKWEPLFEADANMLTLVGDGAIKLNQAIPGFMDADVIRDLTGVKG+D PIP  TEVTT
Sbjct: 361 TVIKWEPLFEADANMLTLVGDGAIKLNQAIPGFMDADVIRDLTGVKGADKPIPAITEVTT 420

Query: 421 DG 422
           DG
Sbjct: 421 DG 422
```

SEQ ID 1452 (GBS364) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 73 (lane 6; MW 50kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 81 (lane 11; MW 75kDa).

GBS364-GST was purified as shown in Figure 216, lane 10.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 452

A DNA sequence (GBSx0489) was identified in *S.agalactiae* <SEQ ID 1455> which encodes the amino acid sequence <SEQ ID 1456>. Analysis of this protein sequence reveals the following:

```
Possible site: 16
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4063(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1457> which encodes the amino acid sequence <SEQ ID 1458>. Analysis of this protein sequence reveals the following:

```
Possible site: 16
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4120(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
```

```
                  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 101/118 (85%), Positives = 110/118 (92%)

Query: 1    MKKKCLICKKTFQAKTNRSLYCSEECRKKGIREKQRKLMKQKRADKKKEKIKVLNTNADV 60
            +KKKCLICKK FQAKTNR+LYCSEECRKKG REKQRKLMKQKRA+++KEK KVLN N DV
Sbjct: 1    LKKKCLICKKNFQAKTNRTLYCSEECRKKGNREKQRKLMKQKRAEQRKEKKKVLNPNTDV 60

Query: 61   TEKPKKIRNLVQHYKKLKREILDNESEFGFTGIALVEGIDIHEENFVDLVMQKIKEQQ 118
            TEKPKKIRNL QHYKKLK+EIL NESEFGFTGI L+EGID+HEENFVDLVMQKIKEQ+
Sbjct: 61   TEKPKKIRNLAQHYKKLKKEILANESEFGFTGITLIEGIDVHEENFVDLVMQKIKEQK 118
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 453

A DNA sequence (GBSx0490) was identified in *S.agalactiae* <SEQ ID 1459> which encodes the amino acid sequence <SEQ ID 1460>. Analysis of this protein sequence reveals the following:

```
Possible site: 19
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0633(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAC39305 GB:AF022773 ORF3 [Lactococcus bacteriophage phi31]
 Identities = 75/109 (68%), Positives = 87/109 (79%), Gaps = 1/109 (0%)

Query: 29   LRADKKGTHRVAFEKNKRRLLKTAHLCGICGRPVDKSLKYPHPLSAAIDHIVPIAKGGHP 88
            LRAD+ G HRVAF+KN++ LLKT + CGICG+P+DK LK P PLS  +DHI+PI KGGHP
Sbjct: 3    LRADRTGAHRVAFDKNRKILLKTQNTCGICGKPIDKRLKAPDPLSPVVDHIIPINKGGHP 62

Query: 89   SSIDNLQLTHWQCNRQKSDKLFINQTAVRATVVGNRNLPQSRDWSSYAS 137
            S++DNLQL HW CNRQKSDKLF N       V+GNRNLPQSRDWSSY S
Sbjct: 63   SAMDNLQLAHWTCNRQKSDKLF-NVKQEEPKVLGNRNLPQSRDWSSYVS 110
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1461> which encodes the amino acid sequence <SEQ ID 1462>. Analysis of this protein sequence reveals the following:

```
Possible site: 49
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4185(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 88/112 (78%), Positives = 102/112 (90%)

Query: 28   KLRADKKGTHRVAFEKNKRRLLKTAHLCGICGRPVDKSLKYPHPLSAAIDHIVPIAKGGH 87
            +LRADKKGTHRVAF++NK++LLK A +CGICG+PVDKSLKYPHPLSAAIDHIVPIAKGGH
Sbjct: 3    QLRADKKGTHRVAFDRNKKKLLKAATVCGICGKPVDKSLKYPHPLSAAIDHIVPIAKGGH 62

Query: 88   PSSIDNLQLTHWQCNRQKSDKLFINQTAVRATVVGNRNLPQSRDWSSYASKE 139
            PS+++NLQLTHWQCNRQKSDKLF NQ +     +GNRNLPQSRDWSS+A K+
Sbjct: 63   PSALENLQLTHWQCNRQKSDKLFANQASNEPKTIGNRNLPQSRDWSSFAFKK 114
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 454

A DNA sequence (GBSx0491) was identified in *S.agalactiae* <SEQ ID 1463> which encodes the amino acid sequence <SEQ ID 1464>. Analysis of this protein sequence reveals the following:

```
Possible site: 33
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4481(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 455

A DNA sequence (GBSx0492) was identified in *S.agalactiae* <SEQ ID 1465> which encodes the amino acid sequence <SEQ ID 1466>. Analysis of this protein sequence reveals the following:

```
Possible site: 28
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2907(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAF43508 GB:AF145054 ORF15 [Streptococcus thermophilus
           bacteriophage 7201]
 Identities = 61/187 (32%), Positives = 90/187 (47%), Gaps = 31/187 (16%)

Query: 1   MNIEEAKKLIDKQSIGKGGVGDIPVVKTHIVKVLLDQIDQPQPEVPRFVADWYEKHKDSL 60
           MN +EA K I K+            +     + L D I   +P VP++VADWYE+HKD
Sbjct: 1   MNRDEAVKKIAKEGY----------ISIEHAEDLYDSIIT-KPVVPQYVADWYEEHKDEF 49

Query: 61  ECDL------YLYHMSIY--DEEVEKDDFYYWMQTSKNPVYTLINMHQFGYTIQKEKLYT 112
             +L      +  H++ Y  +E     DF  W   +KN +  L+NMHQFGY ++KEK YT
Sbjct: 50  YLNLHRVVRDFFEHLNAYYFNENPIDYDFACWYYNTKNAIQILVNMHQFGYEVKKEKRYT 109

Query: 113 VEIPN--PNERQLSFVLMRQLSGNVSIKVMHRDNLDLLKTDNDLQLTESEIRKDFDWAWQ 170
           V I N    E  L++   R+     +     RDN D  +T +    T  E+ ++  + W
Sbjct: 110 VRIRNLDDEETYLNYDKFRE-----TWVFYSRDNTDRFRTIH----THKEL-EEGGFGWV 159

Query: 171 FREEVVE 177
           F  E +E
Sbjct: 160 FDCEGIE 166
```

A related GBS nucleic acid sequence <SEQ ID 10927> which encodes amino acid sequence <SEQ ID 10928> was also identified.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1467> which encodes the amino acid sequence <SEQ ID 1468>. Analysis of this protein sequence reveals the following:

```
Possible site: 21
```

```
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3815(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 70/180 (38%), Positives = 98/180 (53%), Gaps = 30/180 (16%)

Query: 1   MNIEEAKKLIDKQSI-GKGGVGDIPVVKTHIVKVLLDQIDQPQPEVPRFVADWYEKHKDS 59
           MNIEEAK+L+D     GK       V+K   V+ ++DQ++QP+PEVP+ VADW E+ K+
Sbjct: 1   MNIEEAKELVDNSKFYGKTS----SVIKAE-VRDIIDQLNQPKPEVPQCVADWIEECKEE 55

Query: 60  LECDLYLYHMSIYDEEVEKDDFYYWMQTSKNPVYTLINMHQFGYTIQKEKLYTVEIPN-- 117
              DL L    ++         + W+  S      +      GYT++KEKLYTV++PN
Sbjct: 56  ---DLTL--KGLFSNSDMPAKIFDWIFGSDENCRLMAEAWINGYTVEKEKLYTVDLPNGQ 110

Query: 118 PNERQLSFVLMRQLSGNVSIKVMHRDNLDLLKTDNDLQLTESEIRKDFDWAWQFREEVVE 177
           P  R ++ +   Q                 L T+N ++LTESEIRKDF+WAWQF EEV E
Sbjct: 111 PLVRGINTLYFSQN----------------LATEN-VKLTESEIRKDFEWAWQFAEEVTE 153
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 456

A DNA sequence (GBSx0493) was identified in *S.agalactiae* <SEQ ID 1469> which encodes the amino acid sequence <SEQ ID 1470>. Analysis of this protein sequence reveals the following:

```
Possible site: 46
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.5365(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 457

A DNA sequence (GBSx0494) was identified in *S.agalactiae* <SEQ ID 1471> which encodes the amino acid sequence <SEQ ID 1472>. Analysis of this protein sequence reveals the following:

```
Possible site: 50
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -8.55    Transmembrane   34 -  50 (  31 -  54)

----- Final Results -----
               bacterial membrane --- Certainty=0.4418(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9657> which encodes amino acid sequence <SEQ ID 9658> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1473> which encodes the amino acid sequence <SEQ ID 1474>. Analysis of this protein sequence reveals the following:

```
     Possible site: 40
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-11.25   Transmembrane   26 -  42 (  20 -  49)

----- Final Results -----
              bacterial membrane --- Certainty=0.5501(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 56/89 (62%), Positives = 71/89 (78%)

Query: 8  MTEQQMIDCLLYELAKKDKLNIRRNNIITFLSIVLMAISILNVALQDHYKSQITELRTQL 67
           MTE+QMIDCLLYEL KKDK   +++ II  L+++L+ +S L V+L+ +Y+ QI  LRTQL
Sbjct: 1  MTEEQMIDCLLYELVKKDKAIKKKSIIIAALTVMLIVVSGLCVSLKSYYEPQIYGLRTQL 60

Query: 68 SRTQKQLKRASDDRARQTKRIAELTGNGG 96
           SRTQKQLKRAS+   RQTKRIA+LT NGG
Sbjct: 61 SRTQKQLKRASEQNQRQTKRIADLTNNGG 89
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 458

A DNA sequence (GBSx0495) was identified in *S.agalactiae* <SEQ ID 1475> which encodes the amino acid sequence <SEQ ID 1476>. Analysis of this protein sequence reveals the following:

```
Possible site: 32
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2040(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 459

A DNA sequence (GBSx0496) was identified in *S.agalactiae* <SEQ ID 1477> which encodes the amino acid sequence <SEQ ID 1478>. Analysis of this protein sequence reveals the following:

```
Possible site: 34
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3044(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
```

```
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAD37108 GB:AF109874 unknown [Bacteriophage Tuc2009]
 Identities = 50/143 (34%), Positives = 67/143 (45%), Gaps = 29/143 (20%)

Query: 1   MIPNFRAFNKETKKM-YG-VDGFELSVRKIYRCSLADDEFRCGRLETFHFVEDNFDDYIL 58
            MIP  RA++K+ ++M YG V+ F+ S+   YR                HF     +D
Sbjct: 1   MIPKLRAWDKQDERMSYGEVEYFDDSIN--YRFD--------------HFCTGADEDVEF 44

Query: 59  MQSTGMFDKNGVEIFDGDIVLTTRL-------IDY-TYKNFKGVVKMLEGRWLIDTGKDA 110
            MQSTG+ DKNGVEI++GDI+    +       I Y  Y    G   + EG  L    +
Sbjct: 45  MQSTGIKDKNGVEIYEGDILKLHAIFLAPDDKIGYLEYSPKYGYSIICEGNRLY---RQE 101

Query: 111 VGLWTEVDENEAIGNIYQNSELL 133
                T     E IGNIY+N ELL
Sbjct: 102 YWASTNKLNYEVIGNIYENPELL 124
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1479> which encodes the amino acid sequence <SEQ ID 1480>. Analysis of this protein sequence reveals the following:

```
Possible site: 34
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4779(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 44/52 (84%), Positives = 47/52 (89%)

Query: 1  MIPNFRAFNKETKKMYGVDGFELSVRKIYRCSLADDEFRCGRLETFHFVEDN 52
           MIPNFR FNK+TKKMY +DGF+ S RKIYRCSLADDEFR GRLETFHFVEDN
Sbjct: 1  MIPNFRGFNKKTKKMYSIDGFKSSERKIYRCSLADDEFRSGRLETFHFVEDN 52
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 460

A DNA sequence (GBSx0497) was identified in *S.agalactiae* <SEQ ID 1481> which encodes the amino acid sequence <SEQ ID 1482>. Analysis of this protein sequence reveals the following:

```
Possible site: 57
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3843(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9655> which encodes amino acid sequence <SEQ ID 9656> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 461

A DNA sequence (GBSx0498) was identified in *S.agalactiae* <SEQ ID 1483> which encodes the amino acid sequence <SEQ ID 1484>. Analysis of this protein sequence reveals the following:

```
Possible site: 23
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.5189(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9653> which encodes amino acid sequence <SEQ ID 9654> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAF43503 GB:AF145054 ORF10 [Streptococcus thermophilus
           bacteriophage 7201]
 Identities = 92/147 (62%), Positives = 121/147 (81%)

Query: 15   IEPKPQTRPKFSKFGTYEDPKMKRWRKEVSGWIEKNYDGPFFDDCIKVEVTFYMKAPKTL 74
            IEPKPQTRP+FSKFGTYEDPKMK WR+E S  IE+ YDG FF   I V+VTFYMKAP ++
Sbjct: 7    IEPKPQTRPRFSKFGTYEDPKMKAWRRECSRLIEQEYDGQFFYGPISVDVTFYMKAPLSV 66

Query: 75   SKEPTQRSKGKTIQIYQNFVRELIWHAKKPDIDNLIKAVFDSISDAGYDRIQKSGIVWSD 134
            SK+PT +++ KT   ++ F+ E +WH++KPDIDNLIKA+FDSIS AGY+++ K GIVW+D
Sbjct: 67   SKKPTPKARAKTWDAFKKFMAERLWHSRKPDIDNLIKALFDSISTAGYNKVDKKGIVWTD 126

Query: 135 DNIVCDLRAKKKYSQNPRIKVRIEEID 161
            D+IVC L A+K+YS+NPRI+  I+E++
Sbjct: 127 DSIVCKLSAQKRYSENPRIEFEIKELE 153
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 462

A DNA sequence (GBSx0499) was identified in *S.agalactiae* <SEQ ID 1485> which encodes the amino acid sequence <SEQ ID 1486>. Analysis of this protein sequence reveals the following:

```
Possible site: 30
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4007(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 463

A DNA sequence (GBSx0500) was identified in *S.agalactiae* <SEQ ID 1487> which encodes the amino acid sequence <SEQ ID 1488>. This protein is predicted to be pXO1-07. Analysis of this protein sequence reveals the following:

```
Possible site: 26
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3664(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAC38715 GB:AF030367 maturase-related protein [Streptococcus pneumoniae]
 Identities = 146/373 (39%), Positives = 216/373 (57%), Gaps = 18/373 (4%)

Query: 35  LYDKVYRKDILKVAWFYVKRNKGSAGIDDFTIEEIEAYGVQKFLDEIEDQLRNKKYQPKA 94
            L DK+  ++ +  A+  VK NKGSAGID  TIEE++ Y  Q +    ++ ++ +KY+P+
Sbjct: 4   LLDKILSRENMLEAYNQVKSNKGSAGIDGMTIEEMDNYLRQNWR-LTKELIKQRKYKPQP 62

Query: 95  VKRVYIPKANGKKRPLGIPTVRDRVVQTAVKIVIEPIFEADFQEFSYGFRPKRSANQAIR 154
            V +V IPK +G  R LGIPTV DR++Q A+  V+ PI E  F + SYGFRP RS  +AI
Sbjct: 63  VLKVEIPKPDGGIRQLGIPTVMDRMIQQAIVQVMSPICEPHFSDTSYGFRPNRSCEKAIM 122

Query: 155 EIYKYLNYGCEWVIDADLKGYFDTIPHDKLLLLVKERVTDKSIIKLLSLWLEAGIMEDNQ 214
            ++ +YLN G EW++D DL+ +FDT+P D+L+ LV   + D     L+  +L +G++ + Q
Sbjct: 123 KLLEYLNDGYEWIVDIDLEKFFDTVPQDRLMSLVHNIIEDGDTESLIRKYLHSGVIINGQ 182

Query: 215 VRSNILGTPQGGVISPLLANIYLNALDRYWKNNRLEGRGHDAHLIRYADDFVI-LCSNNP 273
                ++GTPQGG +SPLL+NI LN LD+      LE RG     +RYADD VI + S
Sbjct: 183 RYKTLVGTPQGGNLSPLLSNIMLNELDK-----ELEKRG--LRFVRYADDCVITVGSEAA 235

Query: 274 KKYYQYAKQRI--DKLGLTLNEEKTRIVHATEGFDFLGYTLRKSKSHKSGKYKTYYYPSR 331
             K   Y+  R    +LGL +N  KT+I    E   +LG+   KS      +      P +
Sbjct: 236 AKRVMYSVSRFIEKRLGLKVNMTKTKITRPRE-LKYLGFGFWKSSDGWKSR------PHQ 288

Query: 332 KSMKSIKGKVKDVIQTGQHLNLPDVMERLNPMLRGWANYFKAGNSKQHFKSIDNYVIYNL 391
             S++  K K+K + Q    ++L   +E+LN  +RGW NYF  GN K    SID  +   L
Sbjct: 289 DSVRRFKLKLKKLTQRKWSIDLTRRIEQLNLSIRGWINYFSLGNMKSIVASIDERLRTRL 348

Query: 392 TIMLRKKHKKSGK 404
             +++ K+ KK  +
Sbjct: 349 RMIIWKQWKKKSR 361
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 464

A DNA sequence (GBSx0501) was identified in *S.agalactiae* <SEQ ID 1489> which encodes the amino acid sequence <SEQ ID 1490>. Analysis of this protein sequence reveals the following:

```
Possible site: 27
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3833(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9651> which encodes amino acid sequence <SEQ ID 9652> was also identified.

A further related DNA sequence (GBSx2517) was identified in *S.agalactiae* <SEQ ID 7217> which encodes the amino acid sequence <SEQ ID 7218>. Analysis of this protein sequence reveals the following:

```
Possible site: 27
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3833(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1491> which encodes the amino acid sequence <SEQ ID 1492>. Analysis of this protein sequence reveals the following:

```
Possible site: 27
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2299(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 113/163 (69%), Positives = 128/163 (78%), Gaps = 25/163 (15%)

Query: 1   MINNIVLVGRMTKDAELRYTPSNQAVATFSLAVNRNFKNQSGEREADFINCVIWRQQAEN 60
           MINN+VLVGRMTKDAELRYTPS  AVATF+LAVNR FK+Q+GEREADFINCVIWRQ AEN
Sbjct: 1   MINNVVLVGRMTKDAELRYTPSQVAVATFTLAVNRTFKSQNGEREADFINCVIWRQPAEN 60

Query: 61  LANWAKKGALVGITGRIQTRNYENQQGQRIYVTEVVAENFQLLESRNSQQ---------Q 111
           LANWAKKGAL+G+TGRIQTRNYENQQGQR+YVTEVVA+NFQ+LESR +++
Sbjct: 61  LANWAKKGALIGVTGRIQTRNYENQQGQRVYVTEVVADNFQMLESRATREGGSTGSFNGG 120

Query: 112 TNQSGNSSNSY----------------FGNANKMDISDDDLPF 138
            N + +SSNSY                FGN+N MDISDDDLPF
Sbjct: 121 FNNNTSSSNSYSAPAQQTPNFGRDDSPFGNSNPMDISDDDLPF 163
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 465

A DNA sequence (GBSx0502) was identified in *S.agalactiae* <SEQ ID 1493> which encodes the amino acid sequence <SEQ ID 1494>. Analysis of this protein sequence reveals the following:

```
Possible site: 26
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -1.33    Transmembrane   17 -  33 (  17 -  33)

----- Final Results -----
               bacterial membrane --- Certainty=0.1532(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 466

A DNA sequence (GBSx0503) was identified in *S.agalactiae* <SEQ ID 1495> which encodes the amino acid sequence <SEQ ID 1496>. This protein is predicted to be p22 erf-like protein. Analysis of this protein sequence reveals the following:

```
Possible site: 52
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2469(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAA97824 GB:AB044554 orf 17 [Staphylococcus aureus prophage
           phiPV83]
 Identities = 93/183 (50%), Positives = 120/183 (64%), Gaps = 5/183 (2%)

Query: 1   MRKSESITEYAKAFCKAQLEVKQPLKDKDNPFFKSKYVPLENVTEAITTAFANNGISFSQ 60
           M KSE++ E  KA    + EVKQPLKDK+NPFFKSKYVPLENV EAI  A   +G+S++Q
Sbjct: 1   MNKSETVVEINKAMVAFRKEVKQPLKDKNNPFFKSKYVPLENVVEAIDEAATPHGLSYTQ 60

Query: 61  DPTTNTENGYIDVATLVMHTSGEWVEYGPLSVKPTKNDVQGAGSAITYAKRYALSAIFGI 120
               N  +G + VAT++MH SGE++EY P+ +   KN  QGAGS I+Y KRY+LSAIFGI
Sbjct: 61  W-ALNDVDGRVGVATMLMHESGEYIEYDPVFMNAEKNTPQGAGSLISYLKRYSLSAIFGI 119

Query: 121 TSDQDDDGNEDSKPNNSRQSPKATTKKTQKTGYQTPKISNIQIETYKSDLNDIAKATNQN 180
           TSDQDDDGNE S  NN   +PK  T +TQ    +T  I   ++ ++   +    K   QN
Sbjct: 120 TSDQDDDGNEASGKNN---NPKQQT-RTQWASSETIGILRKEVISFTKLIKGTDKEAPQN 175

Query: 181 VEE 183
           + E
Sbjct: 176 IVE 178
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 467

A DNA sequence (GBSx0504) was identified in *S.agalactiae* <SEQ ID 1497> which encodes the amino acid sequence <SEQ ID 1498>. This protein is predicted to be gp157. Analysis of this protein sequence reveals the following:

```
Possible site: 55
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3148(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAD44102 GB:AF115103 orf157 gp [Streptococcus thermophilus
           bacteriophage Sfi21]
```

```
 Identities = 59/160 (36%), Positives = 100/160 (61%), Gaps = 3/160 (1%)

Query: 1    MAYLYELEGIYAQLQSMDLDEETFQDTLDSIDFQSDLENNIEYFVKMLKNVQADAEKYKA 60
            MA LYEL G + ++ +M++D+ET  DTL++ID+ SD EN +E +VK++K+++AD E  K
Sbjct: 1    MATLYELTGQFLEIYNMEIDDETKLDTLEAIDWTSDYENKVEGYVKVIKSLEADIEARKN 60

Query: 61   EKEAFYKKQKQAEAKAEKYKETIRLAMELSQKKKVDAGMFKVSLRRSKKVEILDETKIPL 120
            EK+      K  ++K +K K  + ++M  + + +VD  +FK+   +SK V +++E K+P
Sbjct: 61   EKKRLDGLNKSDQSKIDKLKAALAISMTETGQTRVDTTLFKIGFHKSKAV-VVNEEKLPK 119

Query: 121 DYMQEKIEYKPMKAEISKALKSGIDISGVELIETESLQVK 160
            +Y  +   YKP K  + + LKSG  I G  L E  +L ++
Sbjct: 120 EY--QIATYKPDKKTLKELLKSGKHIEGATLEERRNLNIR 157
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 468

A DNA sequence (GBSx0505) was identified in *S.agalactiae* <SEQ ID 1499> which encodes the amino acid sequence <SEQ ID 1500>. This protein is predicted to be tropomyosin 2. Analysis of this protein sequence reveals the following:

```
Possible site: 26
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4474(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 469

A DNA sequence (GBSx0506) was identified in *S.agalactiae* <SEQ ID 1501> which encodes the amino acid sequence <SEQ ID 1502>. Analysis of this protein sequence reveals the following:

```
Possible site: 21
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4114(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9649> which encodes amino acid sequence <SEQ ID 9650> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 470

A DNA sequence (GBSx0507) was identified in *S.agalactiae* <SEQ ID 1503> which encodes the amino acid sequence <SEQ ID 1504>. Analysis of this protein sequence reveals the following:

```
Possible site: 17
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3799(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1505> which encodes the amino acid sequence <SEQ ID 1506>. Analysis of this protein sequence reveals the following:

```
Possible site: 31
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3775(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 43/46 (93%), Positives = 46/46 (99%)

Query: 1  MTKQHRETLIWYRASHQEREKLLDFGLVDKSQYVTLLRQLRKKYAI 46
           MTKQHRETLIWYRASHQERE+LLDFGLVDK++YVTLLRQLRKKYAI
Sbjct: 1  MTKQHRETLIWYRASHQERERLLDFGLVDKARYVTLLRQLRKKYAI 46
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 471

A DNA sequence (GBSx0508) was identified in *S.agalactiae* <SEQ ID 1507> which encodes the amino acid sequence <SEQ ID 1508>. Analysis of this protein sequence reveals the following:

```
Possible site: 61
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4308(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1509> which encodes the amino acid sequence <SEQ ID 1510>. Analysis of this protein sequence reveals the following:

```
Possible site: 61
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
```

```
                bacterial cytoplasm --- Certainty=0.4308(Affirmative) < succ>
                 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 76/77 (98%), Positives = 76/77 (98%)

Query: 1  MDQEIFNFFNKQIKKDFGKTASKETFAKFASYCAEGIEKNGVKPIFNWINLYAFGTGMTT 60
           MDQEIFNFFNKQIKKDFGKTASKETFAKFASYCAEGIEKNGVKPIFNWINLYAFGTGMTT
Sbjct: 1  MDQEIFNFFNKQIKKDFGKTASKETFAKFASYCAEGIEKNGVKPIFNWINLYAFGTGMTT 60

Query: 61 AEADRLRIERYKQENTL 77
           AEADRLRIERYKQEN L
Sbjct: 61 AEADRLRIERYKQENAL 77
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 472

A DNA sequence (GBSx0509) was identified in *S.agalactiae* <SEQ ID 1511> which encodes the amino acid sequence <SEQ ID 1512>. Analysis of this protein sequence reveals the following:

```
Possible site: 27
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2706(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1513> which encodes the amino acid sequence <SEQ ID 1514>. Analysis of this protein sequence reveals the following:

```
Possible site: 23
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3316(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 52/127 (40%), Positives = 75/127 (58%), Gaps = 1/127 (0%)

Query: 160 EDRFVDVVEANLGRGLVKFEFDMINDYLIGQNVSKDLFLEAVKVAVANNVRKFNYIARIL 219
            E +  +  +   GR +  FE + I  ++   N+  ++   A++ AV NN   + YI +IL
Sbjct: 3   EKKLFENFQLTFGRMISPFEIEDIQKWIHEDNMPIEVVNLALREAVENNKISWKYINKIL 62

Query: 220 DNWINDGIKTPEQAYQAQRDFKAKKANKTMQSQSNVPSWSNPDYKGPDLKEFALGSIDDI 279
             +W   G  T E+     + F   K  +++ + SNVPSWSNPDYK PDL+EFALGS+D I
Sbjct: 63  VDWYKSGDTTVEKVRDRLQRFDDSKKQRSVTT-SNVPSWSNPDYKEPDLEEFALGSMDGI 121

Query: 280 EDGSGDF 286
            EDGSGDF
Sbjct: 122 EDGSGDF 128
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 473

A DNA sequence (GBSx0510) was identified in *S.agalactiae* <SEQ ID 1515> which encodes the amino acid sequence <SEQ ID 1516>. Analysis of this protein sequence reveals the following:

```
Possible site: 26
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -5.63   Transmembrane   13 -  29 (  11 -  31)

----- Final Results -----
              bacterial membrane --- Certainty=0.3251(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9647> which encodes amino acid sequence <SEQ ID 9648> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 474

A DNA sequence (GBSx0511) was identified in *S.agalactiae* <SEQ ID 1517> which encodes the amino acid sequence <SEQ ID 1518>. Analysis of this protein sequence reveals the following:

```
Possible site: 34
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.5822(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 475

A DNA sequence (GBSx0512) was identified in *S.agalactiae* <SEQ ID 1519> which encodes the amino acid sequence <SEQ ID 1520>. Analysis of this protein sequence reveals the following:

```
Possible site: 13
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.4175(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 476

A DNA sequence (GBSx0513) was identified in *S.agalactiae* <SEQ ID 1521> which encodes the amino acid sequence <SEQ ID 1522>. This protein is predicted to be P1-antirepressor homolog. Analysis of this protein sequence reveals the following:

```
Possible site: 37
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3411(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9645> which encodes amino acid sequence <SEQ ID 9646> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAG31333 GB:AF182207 ORF 169a [Bacteriophage mv4]
 Identities = 88/167 (52%), Positives = 122/167 (72%)

Query: 100 MLQRNEKSKQVRKYFIQVEKDFNSPEKIMARALLMADKKITNLTMENNQLQLDLKEAQKQ 159
            M+ +  K K++R+YFIQVEK++NSPE I+ RAL +++ +I  L  +N  L L L+E+ K+
Sbjct: 1   MMSKTAKGKEIRQYFIQVEKNWNSPEMIIQRALEISNARIQELQAQNKSLTLQLEESNKK 60

Query: 160 ARYLDLIIESKGALRVTQIAADYGMSVNKFNKTLLEFGVQHKVNGQWILYKRHMGKGYTD 219
            A YLD+I+ +   L  TQIAADYG S    FN+ L E G+QHKVNGQWILYK +MGKGY
Sbjct: 61  ASYLDIILGTPDLLATTQIAADYGYSARTFNQLLKEVGIQHKVNGQWILYKAYMGKGYVQ 120

Query: 220 SHTFDYQDKNGHTRANVTTTWTQKGRLFLYELLKDNNILPLIEQEDI 266
            S +F ++D+ GH R+  +T WTQKGR  +Y++LK+N  LPLIE++DI
Sbjct: 121 SKSFAFKDRKGHDRSKPSTYWTQKGRKLIYDVLKENGTLPLIERDDI 167
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1523> which encodes the amino acid sequence <SEQ ID 1524>. Analysis of this protein sequence reveals the following:

```
Possible site: 19
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4214(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 130/249 (52%), Positives = 163/249 (65%), Gaps = 14/249 (5%)

Query: 19  MNQLINITLNENQEPVVSGRDLHNVLNIKTQYTKWLERMSEYGFEENVDYIAISQKRLTA 78
            MNQLIN+TLNENQEPVVSGRDLH VL IKTQYTKWLERMSEYGF EN D++AISQKRLTA
Sbjct: 1   MNQLINVTLNENQEPVVSGRDLHKVLEIKTQYTKWLERMSEYGFVENEDFMAISQKRLTA 60

Query: 79  QGNRTEYIDHVLKLDMAKEIAMLQRNEKSKQVRKYFIQVEKDFNSPEKIMARALLMADKK 138
            QGN+TEY DHVLKLDMAKEIAMLQRNEKSK+VRKYFIQVEKDFNSPEKIMARALLMADKK
Sbjct: 61  QGNQTEYTDHVLKLDMAKEIAMLQRNEKSKEVRKYFIQVEKDFNSPEKIMARALLMADKK 120

Query: 139 ITNLTMENNQLQLDLKEAQKQARYLDLIIESKGALRVTQIAA-----DYGMSVNKFNKTL 193
            +       ++L+  ++  + +  + D +  S  ++ V ++A         +   +    L
Sbjct: 121 V-------HKLEAQIEADRPKVLFADAVSASHTSILVGELAKLLKQNGVNIGATRLFTWL 173
```

```
Query: 194 LEFGVQHKVNGQ-WIL-YKRHMGKGYTDSHTFDYQDKNGHTRANVTTTWTQKGRLFLYEL 251
            + G   K NG+ W +  ++ +  G            +GH   + T   T KG+ +
Sbjct: 174 RKHGYLIKRNGRDWNMPTQKSVELGLIRVKETSITHSDGHITVSKTPLVTGKGQQYFINK 233

Query: 252 LKDNNILPL 260
              +   LP+
Sbjct: 234 FLNQEYLPV 242
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 477

A DNA sequence (GBSx0514) was identified in *S.agalactiae* <SEQ ID 1525> which encodes the amino acid sequence <SEQ ID 1526>. Analysis of this protein sequence reveals the following:

```
Possible site: 44
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4205(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1527> which encodes the amino acid sequence <SEQ ID 1528>. Analysis of this protein sequence reveals the following:

```
Possible site: 32
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 21/63 (33%), Positives = 31/63 (48%), Gaps = 1/63 (1%)

Query: 1  MQQFNLKQLREKKGFTQNELADKANVSRSLVVGLETGSYSETSTASLKKLAKALDVKIKD 60
          M+   LK  R  K  +Q  LAD   VSR  +  +E G Y+ T    +  + + LD  + D
Sbjct: 1  MKNLKLKAARAGKDLSQQALADLVGVSRQTIAAVEKGDYNPTINLCI-AICRVLDKTLDD 59

Query: 61 LFF 63
          LF+
Sbjct: 60 LFW 62
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 478

A DNA sequence (GBSx0515) was identified in *S.agalactiae* <SEQ ID 1529> which encodes the amino acid sequence <SEQ ID 1530>. Analysis of this protein sequence reveals the following:

```
Possible site: 26
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0396(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
```

```
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAA17582 GB:D90907 hypothetical protein [Synechocystis sp.]
 Identities = 45/164 (27%), Positives = 79/164 (47%), Gaps = 33/164 (20%)

Query: 102 EEEELRNLFTKLIASSMDKSKNEFNHPSFIEIIKQFDKIDAQNFKIISDLYFKKGFVATG 161
            ++E L+ L+  L+AS++ +S    +  SF+E++KQ D +DA+   ++  L+  +
Sbjct: 97  DDENLQTLWANLLASALTESDRTNSTKSFVEVLKQVDIVDAELLNVLYLLHLRV------ 150

Query: 162 TYYTTIIGQDKPLEHIASHVFVDNLEQNDIAIQSSSLTNLERLGLIQINY--KAHVDEKE 219
                      KP E   ++   D+ + N + I S +L NLERLGL+ I+      VDE+
Sbjct: 151 --------MAKPDEFTYAN---DSRKYNIVQI-SVALNNLERLGLLIIHKYDDTPVDEEA 198

Query: 220 YYNILNNSFITKKNSELKEQNKRVLTNLGMITLTLFGVRFSKTC 263
              +I    ++   N   K            ++LTLFG+ F + C
Sbjct: 199 RISIW---YMQDGNRSFKAH----------VSLTLFGIHFMRVC 229
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1531> which encodes the amino acid sequence <SEQ ID 1532>. Analysis of this protein sequence reveals the following:

```
Possible site: 19
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0151(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 64/215 (29%), Positives = 105/215 (48%), Gaps = 23/215 (10%)

Query: 65  QKLAKEIQDVVSKNIE-NLQEPSLSIAGPALEASKFYLEEEELRNLFTKLIASSMDKSKN 123
            +K   EI    SK  + +L+EP   I  PA+  S+ YL  E LRN+F + IAS+ ++ K
Sbjct: 72  EKFKNEIDCEFSKIPQTSLKEPVEYILYPAINESEQYLSNETLRNMFARTIASTFNQDKE 131

Query: 124 EFNHPSFIEIIKQFDKIDAQNFKIISDLYFKKGFVATGTYYTTIIGQDKPLEHI------ 177
            +  H +F++IIKQ   +DAQN  +I+                  IG     E++
Sbjct: 132 KDLHSAFVQIIKQMTPLDAQNLLLINQ-------EGNNLIANLQIGVHYSKENLSGTVNK 184

Query: 178 ASHVFVDNLEQNDIAIQSSSLTNLERLGLIQINYKAHVDEKEYYNILNNSFITKKNSELK 237
            A+++++  L+ +   I +SS+ NL RLGLI+++Y  +  +  Y +I   +      SE+
Sbjct: 185 ANNIYLSKLDYSPDII-ASSIDNLTRLGLIKVDYLHYPLDSNYESIKQTTIYKSLESEIN 243

Query: 238 EQNKRVLTNL--------GMITLTLFGVRFSKTCL 264
              N    +N         G ++LT FG +F   CL
Sbjct: 244 TLNLFKTSNTKYDIKIEKGKVSLTDFGKKFISVCL 278
```

SEQ ID 1530 (GBS261) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 44 (lane 8; MW 31kDa).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 479

A DNA sequence (GBSx0516) was identified in *S.agalactiae* <SEQ ID 1533> which encodes the amino acid sequence <SEQ ID 1534>. Analysis of this protein sequence reveals the following:

```
Possible site: 16
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -8.55   Transmembrane    3 - 19 (   1 -  26)
```

```
----- Final Results -----
               bacterial membrane --- Certainty=0.4418(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes.*

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 480

A DNA sequence (GBSx0517) was identified in *S.agalactiae* <SEQ ID 1535> which encodes the amino acid sequence <SEQ ID 1536>. Analysis of this protein sequence reveals the following:

```
Possible site: 47
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -4.99    Transmembrane   35 -  51 (  30 -  51)

----- Final Results -----
               bacterial membrane --- Certainty=0.2996(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1537> which encodes the amino acid sequence <SEQ ID 1538>. Analysis of this protein sequence reveals the following:

```
Possible site: 47
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -4.94    Transmembrane   31 -  47 (  30 -  51)

----- Final Results -----
               bacterial membrane --- Certainty=0.2975(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 45/52 (86%), Positives = 48/52 (91%)

Query: 1  MNWKKLMLGDLEHTFTSRDGKEKTSVEFEGGVLPALLVLGGITWLIAWLITK 52
           MNWKKLM GDLEHTFT+ DGKEKTS+EFEGGVLPALLVLGGI W+IAW ITK
Sbjct: 1  MNWKKLMFGDLEHTFTNHDGKEKTSIEFEGGVLPALLVLGGIAWMIAWFITK 52
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 481

A DNA sequence (GBSx0518) was identified in *S.agalactiae* <SEQ ID 1539> which encodes the amino acid sequence <SEQ ID 1540>. Analysis of this protein sequence reveals the following:

```
Possible site: 35
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3445(Affirmative) < succ>
```

```
                 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes.*

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 482

A DNA sequence (GBSx0519) was identified in *S.agalactiae* <SEQ ID 1541> which encodes the amino acid sequence <SEQ ID 1542>. Analysis of this protein sequence reveals the following:

```
   Possible site: 37
   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
                 bacterial cytoplasm --- Certainty=0.3934(Affirmative) < succ>
                  bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                   bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes.*

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 483

A DNA sequence (GBSx0520) was identified in *S.agalactiae* <SEQ ID 1543> which encodes the amino acid sequence <SEQ ID 1544>. This protein is predicted to be repressor protein. Analysis of this protein sequence reveals the following:

```
   Possible site: 61
   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
                 bacterial cytoplasm --- Certainty=0.0905(Affirmative) < succ>
                  bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                   bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9643> which encodes amino acid sequence <SEQ ID 9644> was also identified.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1545> which encodes the amino acid sequence <SEQ ID 1546>. Analysis of this protein sequence reveals the following:

```
   Possible site: 55
   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
                 bacterial cytoplasm --- Certainty=0.3117(Affirmative) < succ>
                  bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                   bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
Identities = 175/264 (66%), Positives = 207/264 (78%), Gaps = 19/264 (7%)
```

```
Query: 34  LGKYIKKYRDTNNLSMAEFAKESGISKAY--VSILEKNRDPRNGKEIIPSIPIIKKVSDT 91
           LG  I+K R+  N++  E ++  G+ K Y  VS  EKN +   GK++       KK+++
Sbjct: 24  LGDRIRKLREGRNMTQTELSEILGM-KTYTTVSKWEKNENFPKGKDL-------KKLAEI 75

Query: 92  IGISFDDLLNSLDENQIVALNETKTEKNLTSSTLQKITSTSSQLEQPRQEKVLSFANEQL 151
             ++ D LL          L ++K  K    +   +I S  +QLEQPRQEKVL+FANEQL
Sbjct: 76  FNVTSDYLLG---------LTDSKLGKITIQNEQPEIVSIYNQLEQPRQEKVLNFANEQL 126

Query: 152 EEQNKVVSMFDRKVEETENYITDYVEGLVAAGLGAYQEDNLHMEVKLRADDVPDKYDTIA 211
           EEQNK VS+FD+K EETE+YITDYVEGLVAAGLGAYQEDNLHM+VKLR+DDVPD+YDTIA
Sbjct: 127 EEQNKTVSIFDKKSEETEDYITDYVEGLVAAGLGAYQEDNLHMKVKLRSDDVPDEYDTIA 186

Query: 212 KVAGNSMEPLIQDNDLLFVKVSSQVDMNDIGIFQVNGKNFVKKLKRDYDGAWYLQSLNKS 271
           KVAG+SMEPLIQDNDLLF+KVSSQVDMNDIGIFQVNGKNFVKKLKRDYDGAWYLQSLNKS
Sbjct: 187 KVAGDSMEPLIQDNDLLFIKVSSQVDMNDIGIFQVNGKNFVKKLKRDYDGAWYLQSLNKS 246

Query: 272 YEEIYLSENDNIRTIGEVVDIYRE 295
           YEEIYLS++D+IRTIGEVVDIYRE
Sbjct: 247 YEEIYLSKDDDIRTIGEVVDIYRE 270
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 484

A DNA sequence (GBSx0521) was identified in *S.agalactiae* <SEQ ID 1547> which encodes the amino acid sequence <SEQ ID 1548>. Analysis of this protein sequence reveals the following:

```
Possible site: 45
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3760(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 485

A DNA sequence (GBSx0522) was identified in *S.agalactiae* <SEQ ID 1549> which encodes the amino acid sequence <SEQ ID 1550>. This protein is predicted to be integrase (ripX). Analysis of this protein sequence reveals the following:

```
Possible site: 20
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2719(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB96616 GB:AJ400629 integrase [Streptococcus pneumoniae
           bacteriophage MM1]
 Identities = 36/59 (61%), Positives = 48/59 (81%), Gaps = 1/59 (1%)
```

```
Query: 2   KIYGDYHTHLFRHSHISFLAEKGIPLNAIMDRVGHSDPKTTLSIYSHTTVNMKE-IINK 59
           KI   +   +H+FRHSHISFLAE G+P+ +IMDRVGHS+ K TL IYSHTT +M++ ++NK
Sbjct: 312 KIEKNLSSHIFRHSHISFLAESGLPIKSIMDRVGHSNAKMTLEIYSHTTEDMEDKLVNK 370
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1551> which encodes the amino acid sequence <SEQ ID 1552>. Analysis of this protein sequence reveals the following:

```
Possible site: 20
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2719(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 63/71 (88%), Positives = 66/71 (92%)

Query: 1  MKIYGDYHTHLFRHSHISFLAEKGIPLNAIMDRVGHSDPKTTLSIYSHTTVNMKEIINKQ 60
           +KIYGDYHTHLFRHSHISFLAEKGIPLNAIMDRVGHSDPKTTLSIYSHTTVNMKEIINKQ
Sbjct: 1  LKIYGDYHTHLFRHSHISFLAEKGIPLNAIMDRVGHSDPKTTLSIYSHTTVNMKEIINKQ 60

Query: 61 TAPFVPLLKSE 71
           T PF   +K +
Sbjct: 61 TDPFKTGIKQK 71
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 486

A DNA sequence (GBSx0523) was identified in *S.agalactiae* <SEQ ID 1553> which encodes the amino acid sequence <SEQ ID 1554>. This protein is predicted to be 50S ribosomal protein L19 (rplS). Analysis of this protein sequence reveals the following:

```
Possible site: 54
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3331(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9641> which encodes amino acid sequence <SEQ ID 9642> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAC01534 GB:U88973 ribosomal protein L19 [Streptococcus thermophilus]
 Identities = 110/115 (95%), Positives = 112/115 (96%)

Query: 25  MNPLIQSLTEGQLRSDIPEFRAGDTVRVHAKVVEGTRERIQIFEGVVISRKGQGISEMYT 84
            MNPLIQSLTEGQLR+DIP FR GDTVRVHAKVVEGTRERIQIFEGVVISRKGQGISEMYT
Sbjct: 1   MNPLIQSLTEGQLRTDIPSFRPGDTVRVHAKVVEGTRERIQIFEGVVISRKGQGISEMYT 60

Query: 85  VRKISGGIGVERTFPIHTPRVDKIEVVRYGKVRRAKLYYLRALQGKAARIKEIRR 139
            VRKIS GIGVERTFPIHTPRVDKIEVVRYGKVRRAKLYYLRALQGKAARIKEIR+
Sbjct: 61  VRKISSGIGVERTFPIHTPRVDKIEVVRYGKVRRAKLYYLRALQGKAARIKEIRK 115
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1555> which encodes the amino acid sequence <SEQ ID 1556>. Analysis of this protein sequence reveals the following:

```
Possible site: 15
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4849(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 111/115 (96%), Positives = 113/115 (97%)

Query: 25  MNPLIQSLTEGQLRSDIPEFRAGDTVRVHAKVVEGTRERIQIFEGVVISRKGQGISEMYT 84
            MNPLIQSLTEGQLRSDIP FR GDTVRVHAKVVEGTRERIQIFEGVVISRKGQGISEMYT
Sbjct: 1   MNPLIQSLTEGQLRSDIPNFRPGDTVRVHAKVVEGTRERIQIFEGVVISRKGQGISEMYT 60

Query: 85  VRKISGGIGVERTFPIHTPRVDKIEVVRYGKVRRAKLYYLRALQGKAARIKEIRR 139
            VRKISGGIGVERTFPIHTPRVDKIEV+R+GKVRRAKLYYLRALQGKAARIKEIRR
Sbjct: 61  VRKISGGIGVERTFPIHTPRVDKIEVIRHGKVRRAKLYYLRALQGKAARIKEIRR 115
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 487

A DNA sequence (GBSx0524) was identified in *S.agalactiae* <SEQ ID 1557> which encodes the amino acid sequence <SEQ ID 1558>. This protein is predicted to be ISL2 protein. Analysis of this protein sequence reveals the following:

```
Possible site: 58
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAC18596 GB:AJ278419 IS1381 transposase [Streptococcus pneumoniae]
 Identities = 111/129 (86%), Positives = 117/129 (90%)

Query: 1   MKAQAIVTSQGRIVSLDIAVNYCHDMKLFKMSRRNIGQAAKILADSGYQGIMKMYSQAQT 60
            MK QAIVTSQGRIVSLDI VNYCHDMKLFKMSRRNIGQA KILADSGYQG+MK+Y QAQT
Sbjct: 1   MKTQAIVTSQGRIVSLDITVNYCHDMKLFKMSRRNIGQAGKILADSGYQGLMKIYPQAQT 60

Query: 61  PRKSSKLKPLTLEDKTYNHTLSKERIKVENIFAKVKTFKIFSTTYRNRRKRFGLRMNLIA 120
             RKSSKLKPLT+EDK  NH LSKER KVENIFAKVKTFK+FSTTYR+ RKRFGLRMNL A
Sbjct: 61  SRKSSKLKPLTVEDKACNHALSKERSKVENIFAKVKTFKMFSTTYRSHRKRFGLRMNLSA 120

Query: 121 GMINRELGF 129
            G+IN ELGF
Sbjct: 121 GIINHELGF 129
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 488

A DNA sequence (GBSx0526) was identified in *S.agalactiae* <SEQ ID 1559> which encodes the amino acid sequence <SEQ ID 1560>. Analysis of this protein sequence reveals the following:

```
Possible site: 61
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-10.99   Transmembrane   81 -  97 (  67 - 107)
   INTEGRAL    Likelihood = -6.32   Transmembrane    8 -  24 (   6 -  25)
   INTEGRAL    Likelihood = -2.76   Transmembrane  120 - 136 ( 120 - 136)

----- Final Results -----
               bacterial membrane --- Certainty=0.5394(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAB04382 GB:AP001509 unknown conserved protein in others
           [Bacillus halodurans]
 Identities = 53/150 (35%), Positives = 82/150 (54%), Gaps = 1/150 (0%)

Query: 1   MLNPYKRIFTLGLLATFLLFIFHFGRYSGLGTNLIEASFTNKNLYDYDWLLKLCLTVITL 60
           M N   R F  GL+   L +I     Y+G G +++E SFT +++  Y +L KL  T +T+
Sbjct: 251 MKNHTVRAFVGGLIIVALTYIIGSYDYNGRGLDMLEDSFT-QDVPPYAFLAKLVFTAVTM 309

Query: 61  AAGYQGGEVTPLFAIGASLGVIIAPILGLPVILVAALGYTSVFGSATNTLLGPILIGGEV 120
             G+ GGE  PLF +GA+LG  +   + LP+  +AALG    FG   NT +   L+G E+
Sbjct: 310 GMGFVGGEAIPLFFVGATLGNTLHAFIDLPLSFLAALGMIVTFGGGANTPIAAFLLGVEM 369

Query: 121 FGFANTPYFVIVCLVAYSISHAHTIYGAQS 150
           F      +F + CL +Y  S  H ++ +Q+
Sbjct: 370 FNGKGIEFFFVACLTSYLFSGHHGLWPSQT 399
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1561> which encodes the amino acid sequence <SEQ ID 1562>. Analysis of this protein sequence reveals the following:

```
     Possible site: 35
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-11.99   Transmembrane   56 -  72 (  53 -  78)
   INTEGRAL    Likelihood = -7.17   Transmembrane  337 - 353 ( 327 - 355)
   INTEGRAL    Likelihood = -6.74   Transmembrane  264 - 280 ( 260 - 282)
   INTEGRAL    Likelihood = -6.16   Transmembrane  167 - 183 ( 161 - 187)
   INTEGRAL    Likelihood = -5.26   Transmembrane  223 - 239 ( 217 - 242)
   INTEGRAL    Likelihood = -5.10   Transmembrane   20 -  36 (  19 -  42)
   INTEGRAL    Likelihood = -0.37   Transmembrane  102 - 118 ( 102 - 119)
   INTEGRAL    Likelihood = -0.16   Transmembrane  300 - 316 ( 300 - 316)

----- Final Results -----
               bacterial membrane --- Certainty=0.5798(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:BAB04382 GB:AP001509 unknown conserved protein in others
           [Bacillus halodurans]
 Identities = 129/397 (32%), Positives = 210/397 (52%), Gaps = 14/397 (3%)

Query: 20  VLGLVGLALPIGGAVGVVDVIFGKGLLFLSEYRDHHLFLLLPFLALAGLVIVFLYDKLG- 78
           +L  +   + IG  VG    +       L E R++  + +L FL LAGL + +LY K G
Sbjct: 9   LLTWIFFGIMIGAIVGSATALLLTVNDHLGETRENRPWFVL-FLPLAGLALGYLYMKAGT 67

Query: 79  ---KEVRQGMGLVFQVGHGQKNQIPPMLIPLILFSTWVTHLFGASAGREGVAVQIGATIS 135
               E+ +G  LV +   G K ++   L PL+   T++T LFG S GREG A+Q+G +++
Sbjct: 68  SAGNELYKGNNLVIESVQG-KGKMLLRLGPLVYLGTFMTILFGGSTGREGAAIQMGGSVA 126
```

```
Query: 136 HYCRR-FVTSQEAARHLLIMGMAAGFAGLFQTPIAAVVFALEVLLVGTLRYSALLPSLVA 194
               + F       R LL+ G++AGF   F TPI A +F +E+  +G L++ AL+P LVA
Sbjct: 127 EAVNKLFKVKLIDTRILLMGGISAGFGAAFGTPITAAIFGMEMASLGRLKFEALVPCLVA 186

Query: 195 AYVASWTSHALG-LEKFTIVLEETLTITPLTLVKLIGLGLIFGLVGNSFAYLL-GWFKPY 252
           ++V  +T+     +E    ++     ++ LT  K+I L ++F LV   +  L  G  K
Sbjct: 187 SFVGHYTTEKFWHVEHEKFIIATVPEVSALTFSKVILLAIVFSLVSVLYCQLRHGIHKLS 246

Query: 253 LSQKLPNPYFRIAFIGALLSICL--MIGHVGRYSGLGTNLIAAAFSGQTILTYDWLLKMI 310
               + N   R AF+G L+ + L  +IG    Y+G G +++  +F+ Q +  Y +L K++
Sbjct: 247 EKYTMKNHTVR-AFVGGLIIVALTYIIGSYD-YNGRGLDMLEDSFT-QDVPPYAFLAKLV 303

Query: 311 VTVISLSAGFQGGEVTPLFAIGASLGIVLAPYLGLPVLLVAALGYTTVFGSATNTFWAPI 370
            T +++  GF GGE  PLF +GA+LG  L  ++ LP+  +AALG    FG   NT  A
Sbjct: 304 FTAVTMGMGFVGGEAIPLFFVGATLGNTLHAFIDLPLSFLAALGMIVTFGGGANTPIAAF 363

Query: 371 FIGIEVFGPENALAYFVTSAAAYMVSHRHSIYSYQKV 407
            +G+E+F  +    +FV    +Y+ S  H ++  Q +
Sbjct: 364 LLGVEMFNGKGIEFFFVACLTSYLFSGHHGLWPSQTI 400
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 91/147 (61%), Positives = 111/147 (74%)

Query: 3   NPYKRIFTLGLLATFLLFIFHFGRYSGLGTNLIEASFTNKNLYDYDWLLKLCLTVITLAA 62
           NPY RI  +G L +  L I H GRYSGLGTNLI A+F+ + +  YDWLLK+ +TVI+L+A
Sbjct: 259 NPYFRIAFIGALLSICLMIGHVGRYSGLGTNLIAAAFSGQTILTYDWLLKMIVTVISLSA 318

Query: 63  GYQGGEVTPLFAIGASLGVIIAPILGLPVILVAALGYTSVFGSATNTLLGPILIGGEVFG 122
           G+QGGEVTPLFAIGASLG+++AP LGLPV+LVAALGYT+VFGSATNT   PI IG EVFG
Sbjct: 319 GFQGGEVTPLFAIGASLGIVLAPYLGLPVLLVAALGYTTVFGSATNTFWAPIFIGIEVFG 378

Query: 123 FANTPYFVIVCLVAYSISHAHTIYGAQ 149
             N   + +    AY +SH H+IY  Q
Sbjct: 379 PENALAYFVTSAAAYMVSHRHSIYSYQ 405
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 489

A DNA sequence (GBSx0527) was identified in *S.agalactiae* <SEQ ID 1563> which encodes the amino acid sequence <SEQ ID 1564>. Analysis of this protein sequence reveals the following:

```
Possible site: 27
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -8.65   Transmembrane   47 -  63 (  45 -  70)
   INTEGRAL    Likelihood = -5.04   Transmembrane  219 - 235 ( 208 - 237)
   INTEGRAL    Likelihood = -3.35   Transmembrane  168 - 184 ( 168 - 187)
   INTEGRAL    Likelihood = -0.48   Transmembrane  141 - 157 ( 141 - 157)

----- Final Results -----
               bacterial membrane --- Certainty=0.4461(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9317> which encodes amino acid sequence <SEQ ID 9318> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAB04382 GB:AP001509 unknown conserved protein in others
           [Bacillus halodurans]
 Identities = 75/223 (33%), Positives = 119/223 (52%), Gaps = 18/223 (8%)
```

```
Query: 17  FSLLIGGVVGAITAVFGRVLLFLTAFRSDYIAYLLPFLSIVGLFIVFVYQKFGGKS---- 72
           F ++IG +VG+ TA+   V   L   R +   ++L FL + GL + ++Y K G  +
Sbjct: 15  FGIMIGAIVGSATALLLTVNDHLGETRENRPWFVL-FLPLAGLALGYLYMKAGTSAGNEL 73

Query: 73  VKGMGLVFEVGHGNEETIPKRLVPLVILTTWLTHLFGGSAGREGVAVQIGATVSHYFQKY 132
            KG  LV E   G  + +  RL PLV L T++T LFGGS GREG A+Q+G +V+    K
Sbjct: 74  YKGNNLVIESVQGKGKML-LRLGPLVYLGTFMTILFGGSTGREGAAIQMGGSVAEAVNKL 132

Query: 133 CRLQNASQLFLVM-GMAAGFAGLFQTPLAATFFAIEVLVVGRLMVSYVLPSLIAALTANF 191
            +++       L+M G++AGF   F TP+ A  F +E+  +GRL    ++P L+A+   ++
Sbjct: 133 FKVKLIDTRILLMGGISAGFGAAFGTPITAAIFGMEMASLGRLKFEALVPCLVASFVGHY 192

Query: 192 VSHSLGLEKFSH------SIATSMALTPDIILKLLVLGLCFGL 228
            +     EKF H       IAT   ++     K+++L + F L
Sbjct: 193 TT-----EKFWHVEHEKFIIATVPEVSALTFSKVILLAIVFSL 230
```

There is also homology to SEQ ID 1562.

A related GBS gene <SEQ ID 8577> and protein <SEQ ID 8578> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 5
McG: Discrim Score:      9.66
GvH: Signal Score (-7.5): -1.12
     Possible site: 27
>>> Seems to have a cleavable N-term signal seq.
ALOM program   count: 7 value: -10.99 threshold:  0.0
   INTEGRAL    Likelihood =-10.99   Transmembrane  328 - 344  ( 314 - 354)
   INTEGRAL    Likelihood = -8.65   Transmembrane   47 -  63  (  45 -  70)
   INTEGRAL    Likelihood = -6.32   Transmembrane  255 - 271  ( 253 - 272)
   INTEGRAL    Likelihood = -4.41   Transmembrane  214 - 230  ( 208 - 238)
   INTEGRAL    Likelihood = -3.35   Transmembrane  168 - 184  ( 168 - 187)
   INTEGRAL    Likelihood = -2.76   Transmembrane  367 - 383  ( 367 - 383)
   INTEGRAL    Likelihood = -0.48   Transmembrane  141 - 157  ( 141 - 157)
   PERIPHERAL  Likelihood =  0.42      94
 modified ALOM score:   2.70

*** Reasoning Step: 3

----- Final Results -----
              bacterial membrane --- Certainty=0.5394(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
ORF01989(349 - 1491 of 1794)
GP|4512350|dbj|BAA75315.1||AB011836(15 - 399 of 424) similar to Bordetella paraperlussis
transposase  for  insertion  sequence  element(27%-identity)  {Bacillus  halodurans}
PIR|T44296|T44296 hypothetical protein [imported] - Bacillus halodurans
%Match = 15.4
%Identity = 33.4  %Similarity = 54.7
Matches = 129  Mismatches = 167  Conservative Sub.s = 82

222       252       282       312       342       372       402       432
MY*RKSKTINLTM*YEQLSKTL*QNLVFIKRRIL*TVIKRFDNYAQYVLVLGFSLLIGGVVGAITAVFGRVLLFLTAFRS
                                                    | ::|| :||: ||::  |   |   |
                                      MNKTFWLTLLTWIFFGIMIGAIVGSATALLLTVNDHLGETRE
                                              10        20        30        40

462       492       513       540       570       600       630       660
DYIAYLLPFLSIVGLFIVFVYQKFG---GKSV-KGMGLVFEVGHGNEETIPKRLVPLVILTTWLTHLFGGSAGREGVAVQ
:   ::| || : || : ::| | |   |:   ||  || |  :| :  :  || ||| | |::| ||||| |||| |:|
NRPWFVL-FLPLAGLALGYLYMKAGTSAGNELYKGNNLVIESVQG-KGKMLLRLGPLVYLGTFMTILFGGSTGREGAAIQ
                   60        70        80         90       100       110       120
```

```
690       720       747       777       807       837       867       894
IGATVSHYFQKYCRLQNASQLFLVMG-MAAGFAGLFQTPLAATFFAIEVLVVGRLMVSYVLPSLIAALTANFVSHSL-GL
:| :|:    |  :::      |:|| ::|||   | ||: |  | :|:  :|||    ::| |:|::  :: :  :  :
MGGSVAEAVNKLFKVKLIDTRILLMGGISAGFGAAFGTPITAAIFGMEMASLGRLKFEALVPCLVASFVGHYTTEKFWHV
       130       140       150       160       170       180       190       200

924       954       984      1014      1041      1071      1101      1131
EKFSHSIATSMALTPDIILKLLVLGLCFGLCGNLFAYLLAKA-KLIASSRLLNPYKRIFTLGLLATFLLFIFHFGRYSGL
|     |||   ::     |:::| : | |   |:  |     ||     : |   | |  ||:   | :|     |:|
EHEKFIIATVPEVSALTFSKVILLAIVFSLVSVLYCQLRHGIHKLSEKYTMKNHTVRAFVGGLIIVALTYIIGSYDYNGR
       210       220       230       240       250       260       270       280

1161      1191      1221      1251      1281      1311      1341      1371
GTNLIEASFTNKNLYDYDWLLKLCLTVITLAAGYQGGEVTPLFAIGASLGVIIAPILGLPVILVAALGYTSVFGSATNTL
| :::| ||| :::  | :| || :| :|:  |: |||  ||| :||:||  :   : ||: ::||||    ||   ||
GLDMLEDSFT-QDVPPYAFLAKLVFTAVTMGMGFVGGEAIPLFFVGATLGNTLHAFIDLPLSFLAALGMIVTFGGGANTP
       290        300       310       320       330       340       350

1401      1431      1461      1491      1521      1551      1581      1611
LGPILIGGEVFGFANTPYFVIVCLVAYSISHAHTIYGAQSR*LVMSFKRVYQFVERNIPFSFLFS*SL*KWSLSIL*MQK
:   |:| |:|      :| : || :|  |  |  :: :|:
IAAFLLGVEMFNGKGIEFFFVACLTSYLFSGHHGLWPSQTIYEPKSRLYGVRKGETIKRTEEMKE
        370       380       390       400       410       420
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 490

A DNA sequence (GBSx0528) was identified in *S.agalactiae* <SEQ ID 1565> which encodes the amino acid sequence <SEQ ID 1566>. Analysis of this protein sequence reveals the following:

```
Possible site: 33
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3568(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAB98234 GB:U67480 chorismate mutase/prephenate dehydratase
          (pheA) [Methanococcus jannaschii]
 Identities = 26/85 (30%), Positives = 46/85 (53%), Gaps = 1/85 (1%)

Query: 2  ELEEIRQEIDEIDQQLVSLLETRMGLILEVIAFKKKHRLPVLDNNRENEVLNNVLKKVQN 61
          +L EIR++IDEID +++ L+   R  L  +V   K +  +P+ D  RE  + + + K  +
Sbjct: 4  KLAEIRKKIDEIDNKILKLIAERNSLAKDVAEIKNQLGIPINDPEREKYIYDRIRKLCKE 63

Query: 62 HQFDDVIRATFKDIMTE-SRVYQKE 85
          H  D+ I      I+ E ++  QK+
Sbjct: 64 HNVDENIGIKIFQILIEHNKALQKQ 88
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1567> which encodes the amino acid sequence <SEQ ID 1568>. Analysis of this protein sequence reveals the following:

```
Possible site: 37
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2356(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 45/91 (49%), Positives = 62/91 (67%)

Query: 1  MELEEIRQEIDEIDQQLVSLLETRMGLILEVIAFKKKHRLPVLDNNRENEVLNNVLKKVQ 60
          M LE+IRQEI+ ID  LV+LLE RM L+ +V A+K  + LPVLD  REN++L+ V   V+
Sbjct: 1  MRLEKIRQEINGIDHHLVALLEKRMALVEQVTAYKLANHLPVLDQARENQILDRVSYLVK 60

Query: 61 NHQFDDVIRATFKDIMTESRVYQKENIVDGD 91
          +  F+  I  TFK IM+ SR YQ +++   GD
Sbjct: 61 DQAFEPAIHETFKTIMSLSRQYQTQHLTGGD 91
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 491

A DNA sequence (GBSx0529) was identified in *S.agalactiae* <SEQ ID 1569> which encodes the amino acid sequence <SEQ ID 1570>. This protein is predicted to be neuraminidase. Analysis of this protein sequence reveals the following:

```
Possible site: 19
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -3.35    Transmembrane   28 -  44 (  28 -  47)

----- Final Results -----
               bacterial membrane --- Certainty=0.2338(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10191> which encodes amino acid sequence <SEQ ID 10192> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAA51473 GB:X72967 neuraminidase [Streptococcus pneumoniae]
 Identities = 294/504 (58%), Positives = 380/504 (75%), Gaps = 10/504 (1%)

Query: 303 EDIKSYFQYYCHLNHQLKLPKGAILSAKTEVYRGGDFGRKNKDNVFGYRIPSLLKTEKGT 362
            E+++   Q +   + + KLP+GA L+ KT+++  G  G+ NKD +  YRIP+LLKT+KGT
Sbjct: 299 EEVQKRSQLFKRSDLEKKLPEGAALTEKTDIFESGRNGKPNKDGIKSYRIPALLKTDKGT 358

Query: 363 LLVGADERIEQACDWGNIGMVIRRSEDDGVTWGKRETIVNLRNNPRVPLVTSGDYSGSPI 422
            L+ GADER   + DWG+IGMVIRRSED+G TWG R TI NLR+NP+     S    GSP+
Sbjct: 359 LIAGADERRLHSSDWGDIGMVIRRSEDNGKTWGDRVTITNLRDNPKA----SDPSIGSPV 414

Query: 423 NMDMALVQDTSSKTKRIFSIYDMFPEGRGVISIANTPEKEYTQIGGQSYLNLYNNGKKSK 482
            N+DM LVQD   +TKRIFSIYDMFPEG+G+  +++  E+ Y +I G++Y  LY  G+K
Sbjct: 415 NIDMVLVQDP--ETKRIFSIYDMFPEGKGIFGMSSQKEEAYKKIDGKTYQILYREGEKG- 471

Query: 483 VFTIRDKGIVYNFKGKKTDYHVITETTKSDHSNLGDIYKGKQLLGNIYFTKHKTSPFRLA 542
             +TIR+ G VY   GK TDY V+ +  K  +S+ GD+YKG QLLGNIYFT +KTSPFR+A
Sbjct: 472 AYTIRENGTVYTPDGKATDYRVVVDPVKPAYSDKGDLYKGNQLLGNIYFTTNKTSPFRIA 531

Query: 543 KSSYVWMSYSDDDGRTWSSPRDITASLRQKGMKFLGIGPGKGIVLKWGPHAGRIIIPAYS 602
            K SY+WMSYSDDDG+TWS+P+DIT  ++   MKFLG+GPG GIVL+ GPH GRI+IP Y+
Sbjct: 532 KDSYLWMSYSDDDGKTWSAPQDITPMVKADWMKFLGVGPGTGIVLRNGPHKGRILIPVYT 591

Query: 603 TNWKSHLRGSQSSRLIYSDDHGKTWHTGKAVNDNRILSNGEKIHSLTMDNKKEQNTESVP 662
            TN  SHL GSQSSR+IYSDDHGKTWH G+AVNDNR + +G+KIHS TM+N++ QNTES
Sbjct: 592 TNNVSHLNGSQSSRIIYSDDHGKTWHAGEAVNDNRQV-DGQKIHSSTMNNRRAQNTESTV 650

Query: 663 VQLKNGDIKLFMRNLTGNLEVATSKDGGETWQNHVKRYKEVHDAYVQLSAIRFEHDKKEY 722
            VQL NGD+KLFMR LTG+L+VATSKDGG TW+  +KRY +V D YVQ+SAI   H+ KEY
Sbjct: 651 VQLNNGDVKLFMRGLTGDLQVATSKDGGVTWEKDIKRYPQVKDVYVQMSAIHTMHEGKEY 710
```

```
Query: 723 ILLVNANGPGKKRQDGYARLAQVNRNGSFKWLYHHHIQDGSFAYNSVQQLNNDKFGVLYE 782
           I+L NA GP  KR++G   LA+V  NG   WL H+ IQ G FAYNS+Q+L N ++G+LYE
Sbjct: 711 IILSNAGGP--KRENGMVHLARVEENGELTWLKHNPIQKGEFAYNSLQELGNGEYGILYE 768

Query: 783 HREKHQNSFTLNYKVFNWSFLSQN 806
           H EK QN++TL+++ FNW FLS++
Sbjct: 769 HTEKGQNAYTLSFRKFNWDFLSKD 792
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 492

A DNA sequence (GBSx0530) was identified in *S.agalactiae* <SEQ ID 1571> which encodes the amino acid sequence <SEQ ID 1572>. This protein is predicted to be unnamed protein product (gatC). Analysis of this protein sequence reveals the following:

```
Possible site: 34
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-12.63   Transmembrane  154 - 170 ( 149 - 178)
   INTEGRAL    Likelihood =-11.99   Transmembrane  103 - 119 (  98 - 123)
   INTEGRAL    Likelihood = -7.91   Transmembrane   21 -  37 (  14 -  40)
   INTEGRAL    Likelihood = -6.53   Transmembrane  448 - 464 ( 444 - 467)
   INTEGRAL    Likelihood = -5.89   Transmembrane   47 -  63 (  45 -  68)
   INTEGRAL    Likelihood = -5.10   Transmembrane  356 - 372 ( 352 - 373)
   INTEGRAL    Likelihood = -4.78   Transmembrane  330 - 346 ( 328 - 350)
   INTEGRAL    Likelihood = -4.41   Transmembrane  376 - 392 ( 375 - 393)
   INTEGRAL    Likelihood = -3.72   Transmembrane  243 - 259 ( 235 - 266)
   INTEGRAL    Likelihood = -2.55   Transmembrane  277 - 293 ( 275 - 293)

----- Final Results -----
            bacterial membrane --- Certainty=0.6052(Affirmative) < succ>
             bacterial outside --- Certainty=0.0000(Not Clear) < succ>
           bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1573> which encodes the amino acid sequence <SEQ ID 1574>. Analysis of this protein sequence reveals the following:

```
     Possible site: 35
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-12.31   Transmembrane  154 - 170 ( 150 - 179)
   INTEGRAL    Likelihood =-11.68   Transmembrane  104 - 120 (  99 - 124)
   INTEGRAL    Likelihood = -9.82   Transmembrane  447 - 463 ( 442 - 469)
   INTEGRAL    Likelihood = -7.91   Transmembrane   22 -  38 (  11 -  41)
   INTEGRAL    Likelihood = -7.11   Transmembrane  377 - 393 ( 375 - 403)
   INTEGRAL    Likelihood = -5.89   Transmembrane   48 -  64 (  46 -  69)
   INTEGRAL    Likelihood = -4.78   Transmembrane  331 - 347 ( 329 - 351)
   INTEGRAL    Likelihood = -3.88   Transmembrane  357 - 373 ( 353 - 373)
   INTEGRAL    Likelihood = -2.55   Transmembrane  278 - 294 ( 276 - 294)
   INTEGRAL    Likelihood = -1.22   Transmembrane  240 - 256 ( 240 - 257)

----- Final Results -----
            bacterial membrane --- Certainty=0.5925(Affirmative) < succ>
             bacterial outside --- Certainty=0.0000(Not Clear) < succ>
           bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 419/482 (86%), Positives = 447/482 (91%)

Query: 1   MQVFLNIVNKFFDPIIHMGSGVVMLIVMTGLAMIFGVKFSKALEGGIKLAIALTGIGAII 60
           MQ FL+I+NK     I +GSGVVMLIVMTGLAMIFGVKF+KALEGGIKLAIALTGIGAII
```

```
Sbjct: 2   MQPFLDIINKILGFPIQLGSGVVMLIVMTGLAMIFGVKFTKALEGGIKLAIALTGIGAII 61

Query: 61  GILTGAFSESLQAFVKNTGINLSIIDVGWAPLATITWGSPYTLYFLLIMLIVNIVMIVMK 120
            GILTGAFSESLQAFVKNTGI+L+IIDVGWAPLATITWGSPYTLYFLL+ML+VNIVMIVMK
Sbjct: 62  GILTGAFSESLQAFVKNTGISLNIIDVGWAPLATITWGSPYTLYFLLVMLVVNIVMIVMK 121

Query: 121 KTDTLDVDIFDIWHLSITGLLIMWYAKKNNLPTLLSVIIATVAIIFVGVLKIINSDLMKP 180
            KTDTLDVDIFDIWHLSITGLLIMWYA +N+LP  +S++IATVA+I VGVLKIINSDLMKP
Sbjct: 122 KTDTLDVDIFDIWHLSITGLLIMWYAARNHLPVFVSLLIATVAVILVGVLKIINSDLMKP 181

Query: 181 TFDDLLGTGPTSPMTSTHMNYMMNPIIMVLDKLFDKVFPGLDKYDFDAAKLNKAIGFWGS 240
            TFDDLLGTGP SPMTSTHMNYMMNPIIMVLDK+FDKVFPGLDKYDFDAAKLNK IGFWGS
Sbjct: 182 TFDDLLGTGPQSPMTSTHMNYMMNPIIMVLDKIFDKVFPGLDKYDFDAAKLNKKIGFWGS 241

Query: 241 KFFIGMILGLVIGIMGNPVFSFAALGGWFSLGFTAGACLELFSLIGSWFIAAVEPLSQGI 300
            KFFIGM LG VIGIMG+P F+  ++  WF LGFTAGACLELFSLIGSWFIAAVEPLSQGI
Sbjct: 242 KFFIGMALGFVIGIMGDPHFTVESIKNWFGLGFTAGACLELFSLIGSWFIAAVEPLSQGI 301

Query: 301 TNFANGKMHGRRFNIGLDWPFIAGRAEIWACANILAPIMLVEAILLSKVGNGILPLAGII 360
            TNFAN +MHGRRFNIGLDWPFIAGRAEIWACANILAPIML+EA+LLSKVGNGILPLAGII
Sbjct: 302 TNFANARMHGRRFNIGLDWPFIAGRAEIWACANILAPIMLIEAVLLSKVGNGILPLAGII 361

Query: 361 AMGVTPALLVVTRGRLIRMITFGTLLLPLFLLSGTMIAPFATELAKKVGAFPAGARAGSL 420
            AMG+TPALLVVTRGRLIRMI FG+LLLPLFLLSGTMIAPFATELAKKVGAFPAG  AGSL
Sbjct: 362 AMGMTPALLVVTRGRLIRMIIFGSLLLPLFLLSGTMIAPFATELAKKVGAFPAGTSAGSL 421

Query: 421 ITHSTLEGPMEKIFGYVIGKATTGQLSAIITLIIFATAYLGLFMWYAKQMKRRNAEYAAN 480
            ITHSTLEGPMEKIFGYVIG+ATTGQ+++IITLIIF   YL LF WYA QMK RNAEYA
Sbjct: 422 ITHSTLEGPMEKIFGYVIGQATTGQIASIITLIIFVAIYLSLFAWYANQMKARNAEYAKT 481

Query: 481 QK 482
             K
Sbjct: 482 MK 483
```

A related GBS gene <SEQ ID 8579> and protein <SEQ ID 8580> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 9
McG: Discrim Score:      4.31
GvH: Signal Score (-7.5): -2.64
     Possible site: 34
>>> Seems to have an uncleavable N-term signal seq
ALOM program   count: 6 value: -12.63 threshold:  0.0
   INTEGRAL    Likelihood =-12.63   Transmembrane  154 - 170 ( 149 - 178)
   INTEGRAL    Likelihood =-11.99   Transmembrane  103 - 119 (  98 - 123)
   INTEGRAL    Likelihood = -7.91   Transmembrane   21 -  37 (  14 -  40)
   INTEGRAL    Likelihood = -5.89   Transmembrane   47 -  63 (  45 -  68)
   INTEGRAL    Likelihood = -4.88   Transmembrane  243 - 259 ( 235 - 265)
   INTEGRAL    Likelihood = -1.22   Transmembrane  268 - 284 ( 268 - 284)
   PERIPHERAL  Likelihood =  0.85      127
 modified ALOM score:   3.03

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.6052(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
ORF00838(343 - 1122 of 1455)
EGAD|91348|EC2092(9 - 344 of 451) PTS system, galactitol specific IIC component
{Escherichia coli} OMNI|NT01EC2494 PTS system galactitol-specific enzyme IIC component
SP|P37189|PTKC_ECOLI PTS SYSTEM, GALACTITOL-SPECIFIC IIC COMPONENT (EIIC-GAT) (GALACTICOL-
PERMEASE IIC COMPONENT) (PHOSPHOTRANSFERASE ENZYME II, C COMPONENT).
GP|1736809|dbj|BAA15955.1||D90847 PTS system, Galactitol-specific IIC component (EIIC-GAT)
```

```
(Galactitol- permease IIC component) (Phosphotransferase enzyme II, C component).
{Escherichia coli} GP|17884
%Match = 10.9
%Identity = 29.8  %Similarity = 59.2
Matches = 68  Mismatches = 88  Conservative Sub.s = 67

282       312       342       372       402       432       462       492
LS*HI*NWN*S*RRRNMQVFLNIVNKFFDPIIHMGSGVVMLIVMTGLAMIFGVKFSKALEGGIKLAIALTGIGAIIGILT
                              |: :|  |:: ||:  :: |:|:|    :: |: : | : ||| :||::
                      MFSEVMRYILDLGPTVMLPIVIIIFSKILGMKAGDCFKAGLHIGIGFVGIGLVIGLML
                              10        20        30        40        50

522       552       582       612       642       672       702
GAFSESLQAFVKNTGINLSIIDVGWAPLATITWGSPYTLYFLLIMLIVNIVMIVMKKTDTLDVDIFDIWHLSITGLLIM-
 :   : :|  :|  :|| ::||||   : :|| |   |  : | ::||: |:: : |  ::|||::|||:: || |:
DSIGPAAKAMAENFDLNLHVVDVGWPGSSPMTWASQIALVAIPIAILVNVAMLLTRMTRVVNVDIWNIWHMTFTGALLHL
          70        80        90       100       110       120       130

747       774       804       834       864       894
------------------------WYAKKN-NLPTLLSVIIATVAIIFVGVLKIINSDLMKPTFDDLLGTGPTSPMTSTH
                        |:|:   |:  |  : |      ::|
ATGSWMIGMAGVVIHAAFVYKLGDWFARDTRNFFELEGIAIPHGTSAYMG------------------------------
         150       160       170       180

924       954       984       1014      1044
MNYMMNPIIMVLDKLFDKVFPGLDKYDFDAAKLNKAIGFWGSKFFIGMILGLVIXIM---------------------~~
      || :::| : :|: ||:::  | |  : :  | :|    :| ::||:| |:
------PIAVLVDAIIEKI-PGVNRIKFSADDIQRKFGPFGEPVTVGFVMGLIIGILAGYDVKGVLQLAVKTAAVML~~~
               200        210       220       230       240       250

1092      1122      1152      1182      1212      1242
~--------------------GNPVFSFASIRWLVFFFVLQQGACLEVGLF*LVSWVQLLQ*NHFLRKLLILLMVNAXX*
                     || |: |  :  : ||  :
~VVSASLIFIPLTILIAVCVPGNQVLPFGDLATIGFFVAMAVAVHRGNLFRTLISGVIIMSITLWIATQTIGLHTQLAAN
             320       330       340       350       360       370       380
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 493

A DNA sequence (GBSx0531) was identified in *S.agalactiae* <SEQ ID 1575> which encodes the amino acid sequence <SEQ ID 1576>. Analysis of this protein sequence reveals the following:

```
Possible site: 14
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0302(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1577> which encodes the amino acid sequence <SEQ ID 1578>. Analysis of this protein sequence reveals the following:

```
Possible site: 14
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0302(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 85/100 (85%), Positives = 96/100 (96%)

Query: 1   MIKILAACGAGVNSSHQIKDAIETQLGDRGYNVHCDAVMVKDITEEMVNKYDIFTPIAKT 60
           MIKILAACGAGVNSSHQIKDAIETQ+ DRGY+VHCDAVMVKDITEE+V++YDIFTPIAKT
Sbjct: 1   MIKILAACGAGVNSSHQIKDAIETQMSDRGYDVHCDAVMVKDITEELVSRYDIFTPIAKT 60

Query: 61  DLGFNVPIPVVEAGPILYRIPVMSEPVFTALEQVIKEHNL 100
           DLGF +PIP+VEAGPILYRIP+MSEPVF  LE+VIKE++L
Sbjct: 61  DLGFEMPIPIVEAGPILYRIPIMSEPVFAELERVIKENHL 100
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 494

A DNA sequence (GBSx0532) was identified in *S.agalactiae* <SEQ ID 1579> which encodes the amino acid sequence <SEQ ID 1580>. This protein is predicted to be GatA. Analysis of this protein sequence reveals the following:

```
Possible site: 15
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2078(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10193> which encodes amino acid sequence <SEQ ID 10194> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAG09977 GB:AF248038 GatA [Streptococcus agalactiae]
 Identities = 39/135 (28%), Positives = 76/135 (55%), Gaps = 9/135 (6%)

Query: 16  QEELFDLVSKALIKQHYVSPNYRQAVKEREREFPTGLKIDLKDGTPIQYVAIPHTETQYC 75
           Q  L +++S+ L+++ YV   + +A+ +RE+++PTGL+++         VAIPHT ++Y
Sbjct: 20  QTNLLEVLSQYLLQKGYVKTEFSKAILQREKDYPTGLQLE------NMAVAIPHTYSEYV 73

Query: 76  LVDRIFYVKNSQPITFKHMINPEEECRVQDFFFIINSRN-SNQSDILSNLITFFITKGNL 134
           L   I+  K  +PI+F  M   E+E  +  +  ++   N  +Q+ +L+ L+T F     +
Sbjct: 74  LKPFIYINKLKEPISFIQM-GTEDEIVMARYVIVLGISNPKDQAGLLAELMTLFSNPKIV 132

Query: 135 DRLHELGDNKEKINH 149
            +L E+   KE + +
Sbjct: 133 QQL-EMAQTKEALKN 146
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1581> which encodes the amino acid sequence <SEQ ID 1582>. Analysis of this protein sequence reveals the following:

```
Possible site: 33
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3130(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 102/154 (66%), Positives = 122/154 (78%)

Query: 4   VTQDILFIDAHSQEELFDLVSKALIKQHYVSPNYRQAVKEREREFPTGLKIDLKDGTPIQ 63
```

```
           V  +ILF +A +Q ELFDLV+  L K  YV+  Y QA+ ERE  FPTGLK+DLKDG+ I
Sbjct: 1   VFPNILFTEARTQPELFDLVASHLEKVGYVTQEYHQALVEREAVFPTGLKVDLKDGSDIL 60

Query: 64  YVAIPHTETQYCLVDRIFYVKNSQPITFKHMINPEEECRVQDFFFIINSRNSNQSDILSN 123
           Y AIPHTET+YCLVD++ YV+NSQ +TFKHMINPEE+C V DFFFIINS+N  Q+ ILSN
Sbjct: 61  YAAIPHTETKYCLVDQVVYVRNSQALTFKHMINPEEDCLVTDFFFIINSQNEGQTTILSN 120

Query: 124 LITFFITKGNLDRLHELGDNKEKINHYLIEKGVF 157
           LITFFITKGNL  L  L D+K+ I++YLIEKGVF
Sbjct: 121 LITFFITKGNLSYLASLKDDKQAISNYLIEKGVF 154
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 495

A DNA sequence (GBSx0533) was identified in *S.agalactiae* <SEQ ID 1583> which encodes the amino acid sequence <SEQ ID 1584>. Analysis of this protein sequence reveals the following:

```
Possible site: 61
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1429(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAA25176 GB:M60447 repressor protein [Lactococcus lactis]
 Identities = 139/255 (54%), Positives = 189/255 (73%), Gaps = 6/255 (2%)

Query: 1   MLKRERLQKIIEKVNINGIVTVNEIMEELDVSDMTVRRDLDELDKAGLLIRIHGGAQKVN 60
           M K+ RL+KI++ + I+G +T+ EI++ELD+SDMT RRDLD L+  GLL R HGGAQ ++
Sbjct: 7   MNKKRRLEKILDMLKIDGTITIKEIIDELDISDMTARRDLDALEADGLLTRTHGGAQLLS 66

Query: 61  ASPTPQNYEKSNTEKYDIQTNEKLEIAQFAKQFINDGETIFIGPGTTLEKLATQLLD--- 117
           +    +  EK++ EK  + T EK++IA+ A   I DG+TIFIGPGTTL +LA +L
Sbjct: 67  SK---KPLEKTHIEKKSLNTKEKIDIAKKACSLIKDGDTIFIGPGTTLVQLALELKGRKG 123

Query: 118 FKIRVVTNSLPVFNILNQSSTLDLILVGGEYREITGAFVGSVTINSIKSLNFSKAFVSSN 177
           +KIRV+TNSLPVF ILN S T+DL+L+GGEYREITGAFVGS+   ++K++ F+KAFV +N
Sbjct: 124 YKIRVITNSLPVFLILNDSETIDLLLLGGEYREITGAFVGSMASTNLKAMRFAKAFVRAN 183

Query: 178 GVFEKSIATYDEGEGEIQRIALNNSFEKFLLVDSQKFGKYDFYTFYQLDDIDFVLTDHNI 237
            V   SIATY + EG IQ++ALNN+ EKFLLVDS KF +YDF+ FY LD +D ++TD+ I
Sbjct: 184 AVTHNSIATYSDKEGVIQQLALNNAVEKFLLVDSTKFDRYDFFNFYNLDQLDTIITDNQI 243

Query: 238 DNVVKEQYSSFTKIL 252
                E++S +T IL
Sbjct: 244 SPQHLEEFSQYTTIL 258
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1585> which encodes the amino acid sequence <SEQ ID 1586>. Analysis of this protein sequence reveals the following:

```
Possible site: 60
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0740(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 161/252 (63%), Positives = 195/252 (76%), Gaps = 3/252 (1%)

Query: 1   MLKRERLQKIIEKVNINGIVTVNEIMEELDVSDMTVRRDLDELDKAGLLIRIHGGAQKVN 60
           MLKRERL KI E VN  GIVTVN+I++ L+VSDMTVRRDLDEL+KAG LIRIHGGAQ +
Sbjct: 1   MLKRERLLKITEIVNEQGIVTVNDIIQTLNVSDMTVRRDLDELEKAGKLIRIHGGAQSIT 60

Query: 61  ASPTPQNYEKSNTEKYDIQTNEKLEIAQFAKQFINDGETIFIGPGTTLEKLATQLLDFKI 120
               P   E+SN EK  +QT EK E+A +A Q +NDGETIFIGPGTTLE  A QL + +I
Sbjct: 61  M---PNKKERSNIEKQTVQTKEKWELASYATQLVNDGETIFIGPGTTLECFAEQLKNRQI 117

Query: 121 RVVTNSLPVFNILNQSSTLDLILVGGEYREITGAFVGSVTINSIKSLNFSKAFVSSNGVF 180
           R+VTNSLPVFNIL  S T+DLIL+GGEYR ITGAFVGS+   +I SL F+KAF+S NG++
Sbjct: 118 RIVTNSLPVFNILQDSETIDLILIGGEYRSITGAFVGSLASQNISSLKFAKAFISCNGIY 177

Query: 181 EKSIATYDEGEGEIQRIALNNSFEKFLLVDSQKFGKYDFYTFYQLDDIDFVLTDHNIDNV 240
           +  IATY E EGEIQ++A NNS EK+LLVD+QKF  YDF+ FY L++ID V+TD  I
Sbjct: 178 KNDIATYSETEGEIQKLAFNNSIEKYLLVDNQKFNAYDFFIFYHLNNIDAVVTDSQITED 237

Query: 241 VKEQYSSFTKIL 252
           V E+YS FT++L
Sbjct: 238 VIERYSQFTQLL 249
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 496

A DNA sequence (GBSx0534) was identified in *S.agalactiae* <SEQ ID 1587> which encodes the amino acid sequence <SEQ ID 1588>. Analysis of this protein sequence reveals the following:

```
Possible site: 19
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3436(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAD13797 GB:AF062533 unknown [Streptococcus agalactiae]
 Identities = 86/371 (23%), Positives = 136/371 (36%), Gaps = 79/371 (21%)

Query: 11  DLSESELKAAQEFLSGKSEANQDKPKTGKTAQEIYEAIEPKAIVKPEDLLFGIAQATDYK 70
           DL++      +  L  K     D   TG         IEP+  V    L      AT
Sbjct: 526 DLTQIAFAEQELMLKDKKHYRYDIVDTG---------IEPRLAVDVSSLPMHAGNATYDT 576

Query: 71  NGTFVIPHKDHYHYVELKWFDEEKDLLADSDKTYSLEDYLATAKYYMMHPEKRPKVEGWG 130
             +FVIPH DH H V   W    +               +AT KY M HPE RP V  W
Sbjct: 577 GSSFVIPHIDHIHVVPYSWLTRNQ---------------IATIKYVMQHPEVRPDV--WS 619

Query: 131 KDAEIYKEKDSNKADKPSPAPTDNKSTSNSSDKNLSAAEVFKQAKPEKIVPLDKIAAHMA 190
           K        + + +  P+  P D ++    +     SA EV      +K +   + AA
Sbjct: 620 KPGH-----EESGSVIPNVTPLDKRAGMPNWQIIHSAEEV------QKALAEGRFAA--- 665

Query: 191 YAVGFEDDQLIVPHHDHYHNVPMAWFDKGGLWKAPEGYTLQQLFST--IKYYMEHPNELP 248
                  D  I    D        W D        +G +L+ +  +   +   +   EL
Sbjct: 666 ------PDGYIFDPRDVLAKETFVWKDGSFSIPRADGSSLRTINKSDLSQAEWQQAQELL 719

Query: 249 KEKGWGHDSDHNKGSNKDNKAKNYAPDEEPEDSGKVTHNYGFYDVNKGSDEEEP-EKQED 307
            +K  G  +D +K            P+E+ +              +K ++ ++P E  ++
Sbjct: 720 AKKNAGDATDTDK------------PEEKQQ-------------ADKSNENQQPSEASKE 754

Query: 308 ESELDEYELGMAQNAKKYGMDRQSFEKQLIQLSNKYSVSFESFNYINGSQVQVTKKDGSK 367
           E E D++      +   YG+DR + E  + QL+ K ++    +       VQ   K+G
Sbjct: 755 EKESDDF----IDSLPDYGLDRATLEDHINQLAQKANID-PKYLIFQPEGVQFYNKNGEL 809
```

```
Query: 368 VLVDIKTLTEV 378
           V  DIKTL ++
Sbjct: 810 VTYDIKTLQQI 820
```

A related DNA sequence was identified in *S.agalactiae* <SEQ ID 6983> which encodes the amino acid sequence <SEQ ID 6984>. Analysis of this protein sequence reveals the following:

```
Possible site: 26
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS gene <SEQ ID 8581> and protein <SEQ ID 8582> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop Possible site: -1   Crend: 2
McG: Discrim Score:      6.06
GvH: Signal Score (-7.5): -5.61
     Possible site: 26
>>> Seems to have an uncleavable N-term signal seq
ALOM program   count: 0 value:   2.23 threshold:  0.0
   PERIPHERAL  Likelihood =  2.23       6
 modified ALOM score:  -0.95

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1589> which encodes the amino acid sequence <SEQ ID 1590>. Analysis of this protein sequence reveals the following:

```
Possible site: 26
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below.

```
 Identities = 808/825 (97%), Positives = 816/825 (97%), Gaps = 3/825 (0%)

Query: 2     KKTYGYIGSVAAILLATHIGSYQLGKHHMGLATKDNQIAYIDDSKGKVKAPKTNKTMDQ 60
             KKTYGYIGSVAAILLATHIGSYQLGKHHMG ATKDNQIAYIDDSKGK KAPKTNKTMDQ
Sbjct: 2     KKTYGYIGSVAAILLATHIGSYQLGKHHMGSATKDNQIAYIDDSKGKAKAPKTNKTMDQ 60

Query: 61   ISAEEGISAEQIVVKITDQGYVTSHGDHYHFYNGKVPYDAIISEELLMTDPNYHFKQSDV 120
            ISAEEGISAEQIVVKITDQGYVTSHGDHYHFYNGKVPYDAIISEELLMTDPNY FKQSDV
Sbjct: 61   ISAEEGISAEQIVVKITDQGYVTSHGDHYHFYNGKVPYDAIISEELLMTDPNYRFKQSDV 120

Query: 121 INEILDGYVIKVNGNYYVYLKPGSKRKNIRTKQQIAEQVAKGTKEAKEKGLAQVAHLSKE 180
            INEILDGYVIKVNGNYYVYLKPGSKRKNIRTKQQIAEQVAKGTKEAKEKGLAQVAHLSKE
Sbjct: 121 INEILDGYVIKVNGNYYVYLKPGSKRKNIRTKQQIAEQVAKGTKEAKEKGLAQVAHLSKE 180

Query: 181 EVAAVNEAKRQGRYTTDDGYIFSPTDIIDDLGDAYLVPHGNHYHYIPKKDLSPSELAAAQ 240
            EVAAVNEAKRQGRYTTDDGYIFSPTDIIDDLGDAYLVPHGNHYHYIPKKDLSPSELAAAQ
Sbjct: 181 EVAAVNEAKRQGRYTTDDGYIFSPTDIIDDLGDAYLVPHGNHYHYIPKKDLSPSELAAAQ 240
```

```
Query: 241 AYWSQKQGRGARPSDYRPTPAP--GRRKAPIPDVTPNPGQGHQPDNGGYHPAPPRPNDAS 298
           AYWSQKQGRGARPSDYRPTPAP  GRRKAPIPDVTPNPGQGHQPDNGGYHPAPPRPNDAS
Sbjct: 241 AYWSQKQGRGARPSDYRPTPAPAPGRRKAPIPDVTPNPGQGHQPDNGGYHPAPPRPNDAS 300

Query: 299 QNKHQRDEFKGKTFKELLDQLHRLDLKYRHVEEDGLIFEPTQVIKSNAFGYVVPHGDHYH 358
           QNKHQRDEFKGKTFKELLDQLHRLDLKYRHVEEDGLIFEPTQVIKSNAFGYVVPHGDHYH
Sbjct: 301 QNKHQRDEFKGKTFKELLDQLHRLDLKYRHVEEDGLIFEPTQVIKSNAFGYVVPHGDHYH 360

Query: 359 IIPRSQLSPLEMELADRYLAGQTDDNDSGSDHSKPSDKEVTHTFLGHRIKAYGKGLDGKP 418
           IIPRSQLSPLEMELADRYLAGQT+D+DSGSDHSKPSDKEVTHTFLGHRIKAYGKGLDGKP
Sbjct: 361 IIPRSQLSPLEMELADRYLAGQTEDDDSGSDHSKPSDKEVTHTFLGHRIKAYGKGLDGKP 420

Query: 419 YDTSDAYVFSKESIHSVDKSGVTAKHGDHFHYIGFGELEQYELDEVANWVKAKGQADELV 478
           YDTSDAYVFSKESIHSVDKSGVTAKHGDHFHYIGFGELEQYELDEVANWVKAKGQADEL
Sbjct: 421 YDTSDAYVFSKESIHSVDKSGVTAKHGDHFHYIGFGELEQYELDEVANWVKAKGQADELA 480

Query: 479 AALDQEQGKEKPLFDTKKVSRKVTKDGKVGYIMPKDGKDYFYARYQLDLTQIAFAEQELM 538
           AALDQEQGKEKPLFDTKKVSRKVTKDGKVGY+MPKDGKDYFYAR QLDLTQIAFAEQELM
Sbjct: 481 AALDQEQGKEKPLFDTKKVSRKVTKDGKVGYMMPKDGKDYFYARDQLDLTQIAFAEQELM 540

Query: 539 LKDKKHYRYDIVDTGIEPRLAVDLSSLPMHAGNATYDTGSSFVIPHIDHIHVVPYSWLTR 598
           LKDKKHYRYDIVDTGIEPRLAVD+SSLPMHAGNATYDTGSSFVIPHIDHIHVVPYSWLTR
Sbjct: 541 LKDKKHYRYDIVDTGIEPRLAVDVSSLPMHAGNATYDTGSSFVIPHIDHIHVVPYSWLTR 600

Query: 599 NQIATIKYVMQHPEVRPDVWSKPGHEESGSVIPNVTPLDKRAGMPNWQIIHSAEEVQKAL 658
           +QIATIKYVMQHPEVRPD+WSKPGHEESGSVIPNVTPLDKRAGMPNWQIIHSAEEVQKAL
Sbjct: 601 DQIATIKYVMQHPEVRPDIWSKPGHEESGSVIPNVTPLDKRAGMPNWQIIHSAEEVQKAL 660

Query: 659 AEGRFAAPDGYIFDPRDVLAKETFVWKDGSFSIPRADGSSLRTINKSDLSQAEWQQAQEL 718
           AEGRFA PDGYIFDPRDVLAKETFVWKDGSFSIPRADGSSLRTINKSDLSQAEWQQAQEL
Sbjct: 661 AEGRFATPDGYIFDPRDVLAKETFVWKDGSFSIPRADGSSLRTINKSDLSQAEWQQAQEL 720

Query: 719 LAKKNAGDATDTDKPEEKQQADKSNENQQPSEASK-EEKESDDFIDSLPDYGLDRATLED 777
           LAKKNAGDATDTDKP+EKQQADKSNENQQPSEASK EEKESDDFIDSLPDYGLDRATLED
Sbjct: 721 LAKKNAGDATDTDKPKEKQQADKSNENQQPSEASKEEEKESDDFIDSLPDYGLDRATLED 780

Query: 778 HINQLAQKANIDPKYLIFQPEGVQFYNKNGELVTYDIKTLQQINP 822
           HINQLAQKANIDPKYLIFQPEGVQFYNKNGELVTYDIKTLQQINP
Sbjct: 781 HINQLAQKANIDPKYLIFQPEGVQFYNKNGELVTYDIKTLQQINP 825
```

SEQ ID 8582 was expressed in *E.coli* in two different forms. GBS293dNterm was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 147 (lane 14; MW 74kDa + lanes 17 & 18; MW 48.8kDa). GBS293C was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figures 148 (lane 2-4; MW 71kDa + lanes 5 & 7; MW 46kDa) and 182 (lane 7; MW 46kDa). Purified GBS293C-His is shown in Figure 241, lanes 8& 9.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 497

A DNA sequence (GBSx0535) was identified in *S.agalactiae* <SEQ ID 1591> which encodes the amino acid sequence <SEQ ID 1592>. Analysis of this protein sequence reveals the following:

```
Possible site: 23
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
               bacterial outside --- Certainty=0.3000(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAD13797 GB:AF062533 unknown [Streptococcus agalactiae]
 Identities = 213/463 (46%), Positives = 277/463 (59%), Gaps = 41/463 (8%)

Query: 4   KKTV-IISALSVALFGTGVGAYQLGSYNA--QKSDNSVSYVKTDKSDSKAQATAVNKTPD 60
           KKT   I +++  L  T +G+YQLG ++      DN ++Y+  D S  K +A   NKT D
Sbjct: 2   KKTYGYIGSVAAILLATHIGSYQLGKHHMGLATKDNQIAYI--DDSKGKVKAPKTNKTMD 59

Query: 61  QISKEEGISAEQIVVKITDDGYVTSHGDHYHYYNGKVPYDAIISEELIMKDPSYVFNKAD 120
           QIS EEGISAEQIVVKITD GYVTSHGDHYH+YNGKVPYDAIISEEL+M DP+Y F ++D
Sbjct: 60  QISAEEGISAEQIVVKITDQGYVTSHGDHYHFYNGKVPYDAIISEELLMTDPNYHFKQSD 119

Query: 121 VINEVKDGYIIKVNGKYYLYLKEGSKRTNVRTKEQIQKQREEWSKGGSKGESGKHSSAKT 180
           VINE+ DGY+IKVNG YY+YLK GSKR N+RTK+QI +Q  + +K     E+ +   A+
Sbjct: 120 VINEILDGYVIKVNGNYYVYLKPGSKRKNIRTKQQIAEQVAKGTK-----EAKEKGLAQV 174

Query: 181 QALS----ASVREAKASGRYTTDDGYVFSPTDVIDDMGDAFLVPHGDHFHYIPKADLSPS 236
             LS    A+V EAK  GRYTTDDGY+FSPTD+IDD+GDA+LVPHG+H+HYIPK DLSPS
Sbjct: 175 AHLSKEEVAAVNEAKRQGRYTTDDGYIFSPTDIIDDLGDAYLVPHGNHYHYIPKKDLSPS 234

Query: 237 ELSAAQAYWNRKTGRSGNSS--KPSNSSSYIHASAPSGNVSTGRHANAPISIPRVTHANH 294
           EL+AAQAYW++K GR    S  +P+ +     A  P    + G+               H
Sbjct: 235 ELAAAQAYWSQKQGRGARPSDYRPTPAPGRRKAPIPDVTPNPGQGHQPD------NGGYH 288

Query: 295 WSKPAGNHATAPKHHAPTTKPINKDSALDKMLKRLYAQPLYARHVESDGLVYDPAQVNAF 354
            + P  N A+  KH     +   K     ++L +L+   L  RHVE DGL+++P QV
Sbjct: 289 PAPPRPNDASQNKHQ----RDEFKGKTFKELLDQLHRLDLKYRHVEEDGLIFEPTQVIKS 344

Query: 355 TAIGVSIPHGNHFHFIHYKDMSPLELE-ATRMVAEHRGHHIDALGKKDSTEKPKHISHEP 413
            A G  +PHG+H+H I    +SPLE+E A R +A          G+ D  +     S
Sbjct: 345 NAFGYVVPHGDHYHIIPRSQLSPLEMELADRYLA----------GQTDDNDSGSDHSKPS 394

Query: 414 NKE-PHTEEEHHAVTPKDQRKGKP---NSQIVYSAQEIEEAKK 452
           +KE  HT   H          GKP   +   V+S + I    K
Sbjct: 395 DKEVTHTFLGHRIKAYGKGLDGKPYDTSDAYVFSKESIHSVDK 437
```

There is also homology to SEQ ID 1590.

SEQ ID 1592 (GBS94) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 17 (lane 3; MW 52.5kDa).

GBS94-His was purified as shown in Figure 194, lane 8.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 498

A DNA sequence (GBSx0536) was identified in *S.agalactiae* <SEQ ID 1593> which encodes the amino acid sequence <SEQ ID 1594>. This protein is predicted to be Lmb. Analysis of this protein sequence reveals the following:

```
Possible site: 24
>>> May be a lipoprotein

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

There is also homology to SEQ IDs 1596 and 5548.

A related GBS gene <SEQ ID 8583> and protein <SEQ ID 8584> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: 22   Crend: 5
McG: Discrim Score:     13.64
GvH: Signal Score (-7.5): -5.75
     Possible site: 24
>>> May be a lipoprotein
ALOM program   count: 0 value:   4.83 threshold:  0.0
   PERIPHERAL  Likelihood =  4.83      33
 modified ALOM score:  -1.47

----- Final Results -----
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

SEQ ID 8584 (GBS22) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 14 (lane 6; MW 35kDa).

The GBS22-His fusion product was purified (Figure 94A; see also Figure 193, lane 4) and used to immunise mice (lane 2 product; 20μg/mouse). The resulting antiserum was used for Western blot (Figure 94B), FACS (Figure 94C), and in the *in vivo* passive protection assay (Table III). These tests confirm that the protein is immunoaccessible on GBS bacteria and that it is an effective protective immunogen.

SEQ ID 8584 (GBS22) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 183 (lane 7 & 8; MW 35kDa).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 499

A DNA sequence (GBSx0537) was identified in *S.agalactiae* <SEQ ID 1597> which encodes the amino acid sequence <SEQ ID 1598>. Analysis of this protein sequence reveals the following:

```
Possible site: 39
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.59   Transmembrane   19 -  35 (  19 -  35)

----- Final Results -----
              bacterial membrane --- Certainty=0.1235(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAA51352 GB:X72832 ORF1 [Streptococcus equisimilis]
 Identities = 104/145 (71%), Positives = 126/145 (86%)

Query: 1   MKIIIQRVNQASVSIEDDVVGSIEKGLVLLVGIAPEDTTEDIAYAVRKITSMRIFSDDEG 60
            MK+++QRV +ASVSI+  + G+I +GL+LLVG+ P+D  ED+AYAVRKI +MRIFSD +G
Sbjct: 1   MKLVLQRVKEASVSIDGKIAGAINQGLLLLVGVGPDDAAEDLAYAVRKIVNMRIFSDADG 60

Query: 61  KMNLSIQDIKGSVLSISQFTLFADTKKGNRPAFTGAADPVKANQFYDIFNQELANHVSVE 120
            KMN SIQDIKGS+LS+SQFTL+ADTKKGNRPAFTGAA P  A+QFYD FN++LA+ V VE
Sbjct: 61  KMNQSIQDIKGSILSVSQFTLYADTKKGNRPAFTGAAKPDMASQFYDRFNEQLADFVPVE 120

Query: 121 TGQFGADMQVSLINDGPVTIVLDTK 145
             G FGADMQVSLINDGPVTI+LDTK
Sbjct: 121 RGVFGADMQVSLINDGPVTIILDTK 145
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1599> which encodes the amino acid sequence <SEQ ID 1600>. Analysis of this protein sequence reveals the following:

```
Possible site: 39
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1430(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 103/145 (71%), Positives = 124/145 (85%)

Query: 1   MKIIIQRVNQASVSIEDDVVGSIEKGLVLLVGIAPEDTTEDIAYAVRKITSMRIFSDDEG 60
           MK+++QRV +ASVSI+  + G+I +GL+LLVG+ P+D  ED+AYAVRKI +MRIFSD +G
Sbjct: 1   MKLVLQRVKEASVSIDGKIAGAINQGLLLLVGVGPDDNAEDLAYAVRKIVNMRIFSDADG 60

Query: 61  KMNLSIQDIKGSVLSISQFTLFADTKKGNRPAFTGAADPVKANQFYDIFNQELANHVSVE 120
           KMN SIQDIKGS+LS+SQFTL+ADTKKGNRPAFTGAA P  A+Q YD FN++LA  V VE
Sbjct: 61  KMNQSIQDIKGSILSVSQFTLYADTKKGNRPAFTGAAKPDLASQLYDSFNEQLAEFVPVE 120

Query: 121 TGQFGADMQVSLINDGPVTIVLDTK 145
            G FGADMQVSLINDGPVTI+LDTK
Sbjct: 121 RGVFGADMQVSLINDGPVTIILDTK 145
```

SEQ ID 1598 (GBS368) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 64 (lane 4; MW 20kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 70 (lane 4; MW 45kDa).

GBS368-GST was purified as shown in Figure 215, lane 6.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 500

A DNA sequence (GBSx0538) was identified in *S.agalactiae* <SEQ ID 1601> which encodes the amino acid sequence <SEQ ID 1602>. This protein is predicted to be stringent response-like protein (rel) (relA). Analysis of this protein sequence reveals the following:

```
Possible site: 37
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.32   Transmembrane   60 -  76 (  60 -  76)

----- Final Results -----
               bacterial membrane --- Certainty=0.1128(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAA51353 GB:X72832 stringent response-like protein
           [Streptococcus equisimilis]
 Identities = 647/739 (87%), Positives = 696/739 (93%), Gaps = 1/739 (0%)

Query: 1   MVKEINLTGEEVVAITSQYMSETDVAFVKFALNYATAAHYYQARKSGEPYIIHPIQVAGI 60
           M KEINLTGEEVVA+ ++YM+ETD AFVK AL+YATAAH+YQ RKSGEPYI+HPIQVAGI
Sbjct: 1   MAKEINLTGEEVVALAAKYMNETDAAFVKKALDYATAAHFYQVRKSGEPYIVHPIQVAGI 60
```

```
Query: 61  LADLHLDAVTVACGFLHDVVEDTEITLDEIETDFGKDVRDIIDGVTKLGKVEYKSHEEQL 120
           LADLHLDAVTVACGFLHDVVEDT+ITLD IE DFGKDVRDI+DGVTKLGKVEYKSHEEQL
Sbjct: 61  LADLHLDAVTVACGFLHDVVEDTDITLDNIEFDFGKDVRDIVDGVTKLGKVEYKSHEEQL 120

Query: 121 AENHRKMLMAMSKDIRVILVKLADRLHNMRTLKHLRKDKQERISRETMEIYAPLAHRLGI 180
           AENHRKMLMAMSKDIRVILVKLADRLHNMRTLKHLRKDKQERISRETMEIYAPLAHRLGI
Sbjct: 121 AENHRKMLMAMSKDIRVILVKLADRLHNMRTLKHLRKDKQERISRETMEIYAPLAHRLGI 180

Query: 181 SRIKWELEDLSFRYLNETEFYKISHMMSEKRREREELVDIIVDKIRSYTEEQGLYGDIYG 240
           SRIKWELEDL+FRYLNETEFYKISHMM+EKRRERE LVD IV KI+SYT EQGL+GD+YG
Sbjct: 181 SRIKWELEDLAFRYLNETEFYKISHMMNEKRREREALVDDIVTKIKSYTTEQGLFGDVYG 240

Query: 241 RPKHIYSIYRKMRDKKKRFDQIYDLIAIRCIMETASDVYAMVGYIHELWRPMPGRFKDYI 300
           RPKHIYSIYRKMRDKKKRFDQI+DLIAIRC+MET SDVYAMVGYIHELWRPMPGRFKDYI
Sbjct: 241 RPKHIYSIYRKMRDKKKRFDQIFDLIAIRCVMETQSDVYAMVGYIHELWRPMPGRFKDYI 300

Query: 301 AAPKANGYQSIHTTVYGPKGPIEIQIRTKEMHQVAEFGVAAHWAYKKGITSKVNQAEQSV 360
           AAPKANGYQSIHTTVYGPKGPIEIQIRTKEMHQVAE+GVAAHWAYKKG+   KVNQAEQ V
Sbjct: 301 AAPKANGYQSIHTTVYGPKGPIEIQIRTKEMHQVAEYGVAAHWAYKKGVRGKVNQAEQKV 360

Query: 361 GMGWIQELVELQDESK-DAKDFVDSVKEDIFTERIYVFTPNGAVQELPRESGPIDFAYAI 419
           GM WI+ELVELQD S  DA DFVDSVKEDIF+ERIYVFTP GAVQELP++SGPIDFAYAI
Sbjct: 361 GMNWIKELVELQDASNGDAVDFVDSVKEDIFSERIYVFTPTGAVQELPKDSGPIDFAYAI 420

Query: 420 HTQVGEKATGAKVNGRMVPLTAKLKTGDVVEIITNPNSFGPSRDWIKIVKTNKARNKIRQ 479
           HTQVGEKA GAKVNGRMVPLTAKLKTGDVVEI+TNPNSFGPSRDWIK+VKTNKARNKIRQ
Sbjct: 421 HTQVGEKAIGAKVNGRMVPLTAKLKTGDVVEIVTNPNSFGPSRDWIKLVKTNKARNKIRQ 480

Query: 480 FFKNQDKETSINKGRELLVDYFQEQGYVPNKYLDKKHIEEILPRVSVKSEEALYAAVGFG 539
           FFKNQDKE S+NKGR++LV YFQEQGYV NKYLDKK IE ILP+VSVKSEE+LYAAVGFG
Sbjct: 481 FFKNQDKELSVNKGRDMLVSYFQEQGYVANKYLDKKRIEAILPKVSVKSEESLYAAVGFG 540

Query: 540 DLSPISIFNKLTEKERREEERAKAKAEADELINGGEIKTDKRDVLKVKSENGVIIQGASG 599
           D+SP+S+FNKLTEKERREEERAKAKAEA+EL+NGGEIK + +DVLKV+SENGVIIQGASG
Sbjct: 541 DISPVSVFNKLTEKERREEERAKAKAEAEELVNGGEIKHENKDVLKVRSENGVIIQGASG 600

Query: 600 LLMRIAKCCNPVPGDLIEGYITKGRGVAIHRSDCQNLKSQENYEQRLIDVEWDDDGSKKE 659
           LLMRIAKCCNPVPGD IEGYITKGRG+AIHR+DC N+KSQ+ Y++RLI+VEWD D S K+
Sbjct: 601 LLMRIAKCCNPVPGDPIEGYITKGRGIAIHRADCNNIKSQDGYQERLIEVEWDLDNSSKD 660

Query: 660 YMAEIDIYGLNRSGLLNDVLQTLSNATKLVSTVNAQPTKDMKFANIHVSFGISNLAQLTT 719
           Y AEIDIYGLNR GLLNDVLQ LSN+TK +STVNAQPTKDMKFANIHVSFGI NL  LTT
Sbjct: 661 YQAEIDIYGLNRRGLLNDVLQILSNSTKSISTVNAQPTKDMKFANIHVSFGIPNLTHLTT 720

Query: 720 VVDKIKIIPDVYSVKRTNG 738
           VV+KIK +PDVYSVKRTNG
Sbjct: 721 VVEKIKAVPDVYSVKRTNG 739
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1603> which encodes the amino acid sequence <SEQ ID 1604>. Analysis of this protein sequence reveals the following:

```
        Possible site: 41
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.32   Transmembrane   64 -  80 (  64 -  80)

----- Final Results -----
               bacterial membrane --- Certainty=0.1128(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:CAA51353 GB:X72832 stringent response-like protein
          [Streptococcus equisimilis]
 Identities = 700/739 (94%), Positives = 721/739 (96%)

Query: 5   MAKIMNVTGEEVIALAATYMTKADVAFVAKALAYATAAHFYQVRKSGEPYIVHPIQVAGI 64
```

```
           MAK +N+TGEEV+ALAA YM + D AFV KAL YATAAHFYQVRKSGEPYIVHPIQVAGI
Sbjct: 1   MAKEINLTGEEVVALAAKYMNETDAAFVKKALDYATAAHFYQVRKSGEPYIVHPIQVAGI 60

Query: 65  LADLHLDAVTVACGFLHDVVEDTDITLDEIEADFGHDARDIVDGVTKLGEVEYKSHEEQL 124
           LADLHLDAVTVACGFLHDVVEDTDITLD IE DFG D RDIVDGVTKLG+VEYKSHEEQL
Sbjct: 61  LADLHLDAVTVACGFLHDVVEDTDITLDNIEFDFGKDVRDIVDGVTKLGKVEYKSHEEQL 120

Query: 125 AENHRKMLMAMSKDIRVILVKLADRLHNMRTLKHLRKDKQERISRETMEIYAPLAHRLGI 184
           AENHRKMLMAMSKDIRVILVKLADRLHNMRTLKHLRKDKQERISRETMEIYAPLAHRLGI
Sbjct: 121 AENHRKMLMAMSKDIRVILVKLADRLHNMRTLKHLRKDKQERISRETMEIYAPLAHRLGI 180

Query: 185 SRIKWELEDLAFRYLNETEFYKISHMMKEKRREREALVEAIVSKVKTYTTQQGLFGDVYG 244
           SRIKWELEDLAFRYLNETEFYKISHMM EKRREREALV+ IV+K+K+YTT+QGLFGDVYG
Sbjct: 181 SRIKWELEDLAFRYLNETEFYKISHMMNEKRREREALVDDIVTKIKSYTTEQGLFGDVYG 240

Query: 245 RPKHIYSIYRKMRDKKKRFDQIFDLIAIRCVMETQSDVYAMVGYIHELWRPMPGRFKDYI 304
           RPKHIYSIYRKMRDKKKRFDQIFDLIAIRCVMETQSDVYAMVGYIHELWRPMPGRFKDYI
Sbjct: 241 RPKHIYSIYRKMRDKKKRFDQIFDLIAIRCVMETQSDVYAMVGYIHELWRPMPGRFKDYI 300

Query: 305 AAPKANGYQSIHTTVYGPKGPIEIQIRTKDMHQVAEYGVAAHWAYKKGVRGKVNQAEQAV 364
           AAPKANGYQSIHTTVYGPKGPIEIQIRTK+MHQVAEYGVAAHWAYKKGVRGKVNQAEQ V
Sbjct: 301 AAPKANGYQSIHTTVYGPKGPIEIQIRTKEMHQVAEYGVAAHWAYKKGVRGKVNQAEQKV 360

Query: 365 GMNWIKELVELQDASNGDAVDFVDSVKEDIFSERIYVFTPTGAVQELPKESGPIDFAYAI 424
           GMNWIKELVELQDASNGDAVDFVDSVKEDIFSERIYVFTPTGAVQELPK+SGPIDFAYAI
Sbjct: 361 GMNWIKELVELQDASNGDAVDFVDSVKEDIFSERIYVFTPTGAVQELPKDSGPIDFAYAI 420

Query: 425 HTQIGEKATGAKVNGRMVPLTAKLKTGDVVEIITNANSFGPSRDWVKLVKTNKARNKIRQ 484
           HTQ+GEKA GAKVNGRMVPLTAKLKTGDVVEI+TN NSFGPSRDW+KLVKTNKARNKIRQ
Sbjct: 421 HTQVGEKAIGAKVNGRMVPLTAKLKTGDVVEIVTNPNSFGPSRDWIKLVKTNKARNKIRQ 480

Query: 485 FFKNQDKELSVNKGRDLLVSYFQEQGYVANKYLDKKRIEAILPKVSVKSEESLYAAVGFG 544
           FFKNQDKELSVNKGRD+LVSYFQEQGYVANKYLDKKRIEAILPKVSVKSEESLYAAVGFG
Sbjct: 481 FFKNQDKELSVNKGRDMLVSYFQEQGYVANKYLDKKRIEAILPKVSVKSEESLYAAVGFG 540

Query: 545 DISPISVFNKLTEKERREEERAKAKAEAEELVKGGEVKHENKDVLKVRSENGVIIQGASG 604
           DISP+SVFNKLTEKERREEERAKAKAEAEELV GGE+KHENKDVLKVRSENGVIIQGASG
Sbjct: 541 DISPVSVFNKLTEKERREEERAKAKAEAEELVNGGEIKHENKDVLKVRSENGVIIQGASG 600

Query: 605 LLMRIAKCCNPVPGDPIDGYITKGRGIAIHRSDCHNIKSQDGYQERLIEVEWDLDNSSKD 664
           LLMRIAKCCNPVPGDPI+GYITKGRGIAIHR+DC+NIKSQDGYQERLIEVEWDLDNSSKD
Sbjct: 601 LLMRIAKCCNPVPGDPIEGYITKGRGIAIHRADCNNIKSQDGYQERLIEVEWDLDNSSKD 660

Query: 665 YQAEIDIYGLNRSGLLNDVLQILSNSTKSISTVNAQPTKDMKFANIHVSFGIPNLTHLTT 724
           YQAEIDIYGLNR GLLNDVLQILSNSTKSISTVNAQPTKDMKFANIHVSFGIPNLTHLTT
Sbjct: 661 YQAEIDIYGLNRRGLLNDVLQILSNSTKSISTVNAQPTKDMKFANIHVSFGIPNLTHLTT 720

Query: 725 VVEKIKAVPDVYSVKRTNG 743
           VVEKIKAVPDVYSVKRTNG
Sbjct: 721 VVEKIKAVPDVYSVKRTNG 739
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 635/739 (85%), Positives = 691/739 (92%), Gaps = 1/739 (0%)

Query: 1   MVKEINLTGEEVVAITSQYMSETDVAFVKFALNYATAAHYYQARKSGEPYIIHPIQVAGI 60
           M K +N+TGEEV+A+ + YM++ DVAFV  AL YATAAH+YQ RKSGEPYI+HPIQVAGI
Sbjct: 5   MAKIMNVTGEEVIALAATYMTKADVAFVAKALAYATAAHFYQVRKSGEPYIVHPIQVAGI 64

Query: 61  LADLHLDAVTVACGFLHDVVEDTEITLDEIETDFGKDVRDIIDGVTKLGKVEYKSHEEQL 120
           LADLHLDAVTVACGFLHDVVEDT+ITLDEIE DFG D RDI+DGVTKLG+VEYKSHEEQL
Sbjct: 65  LADLHLDAVTVACGFLHDVVEDTDITLDEIEADFGHDARDIVDGVTKLGEVEYKSHEEQL 124

Query: 121 AENHRKMLMAMSKDIRVILVKLADRLHNMRTLKHLRKDKQERISRETMEIYAPLAHRLGI 180
           AENHRKMLMAMSKDIRVILVKLADRLHNMRTLKHLRKDKQERISRETMEIYAPLAHRLGI
Sbjct: 125 AENHRKMLMAMSKDIRVILVKLADRLHNMRTLKHLRKDKQERISRETMEIYAPLAHRLGI 184

Query: 181 SRIKWELEDLSFRYLNETEFYKISHMMSEKRREREELVDIIVDKIRSYTEEQGLYGDIYG 240
```

```
           SRIKWELEDL+FRYLNETEFYKISHMM EKRRERE LV+ IV K+++YT +QGL+GD+YG
Sbjct: 185 SRIKWELEDLAFRYLNETEFYKISHMMKEKRREREALVEAIVSKVKTYTTQQGLFGDVYG 244

Query: 241 RPKHIYSIYRKMRDKKKRFDQIYDLIAIRCIMETASDVYAMVGYIHELWRPMPGRFKDYI 300
           RPKHIYSIYRKMRDKKKRFDQI+DLIAIRC+MET SDVYAMVGYIHELWRPMPGRFKDYI
Sbjct: 245 RPKHIYSIYRKMRDKKKRFDQIFDLIAIRCVMETQSDVYAMVGYIHELWRPMPGRFKDYI 304

Query: 301 AAPKANGYQSIHTTVYGPKGPIEIQIRTKEMHQVAEFGVAAHWAYKKGITSKVNQAEQSV 360
           AAPKANGYQSIHTTVYGPKGPIEIQIRTK+MHQVAE+GVAAHWAYKKG+  KVNQAEQ+V
Sbjct: 305 AAPKANGYQSIHTTVYGPKGPIEIQIRTKDMHQVAEYGVAAHWAYKKGVRGKVNQAEQAV 364

Query: 361 GMGWIQELVELQDESK-DAKDFVDSVKEDIFTERIYVFTPNGAVQELPRESGPIDFAYAI 419
           GM WI+ELVELQD S  DA DFVDSVKEDIF+ERIYVFTP GAVQELP+ESGPIDFAYAI
Sbjct: 365 GMNWIKELVELQDASNGDAVDFVDSVKEDIFSERIYVFTPTGAVQELPKESGPIDFAYAI 424

Query: 420 HTQVGEKATGAKVNGRMVPLTAKLKTGDVVEIITNPNSFGPSRDWIKIVKTNKARNKIRQ 479
           HTQ+GEKATGAKVNGRMVPLTAKLKTGDVVEIITN NSFGPSRDW+K+VKTNKARNKIRQ
Sbjct: 425 HTQIGEKATGAKVNGRMVPLTAKLKTGDVVEIITNANSFGPSRDWVKLVKTNKARNKIRQ 484

Query: 480 FFKNQDKETSINKGRELLVDYFQEQGYVPNKYLDKKHIEEILPRVSVKSEEALYAAVGFG 539
           FFKNQDKE S+NKGR+LLV YFQEQGYV NKYLDKK IE ILP+VSVKSEE+LYAAVGFG
Sbjct: 485 FFKNQDKELSVNKGRDLLVSYFQEQGYVANKYLDKKRIEAILPKVSVKSEESLYAAVGFG 544

Query: 540 DLSPISIFNKLTEKERREEERAKAKAEADELINGGEIKTDKRDVLKVKSENGVIIQGASG 599
           D+SPIS+FNKLTEKERREEERAKAKAEA+EL+ GGE+K + +DVLKV+SENGVIIQGASG
Sbjct: 545 DISPISVFNKLTEKERREEERAKAKAEAEELVKGGEVKHENKDVLKVRSENGVIIQGASG 604

Query: 600 LLMRIAKCCNPVPGDLIEGYITKGRGVAIHRSDCQNLKSQENYEQRLIDVEWDDDGSKKE 659
           LLMRIAKCCNPVPGD I+GYITKGRG+AIHRSDC N+KSQ+ Y++RLI+VEWD D S K+
Sbjct: 605 LLMRIAKCCNPVPGDPIDGYITKGRGIAIHRSDCHNIKSQDGYQERLIEVEWDLDNSSKD 664

Query: 660 YMAEIDIYGLNRSGLLNDVLQTLSNATKLVSTVNAQPTKDMKFANIHVSFGISNLAQLTT 719
           Y AEIDIYGLNRSGLLNDVLQ LSN+TK +STVNAQPTKDMKFANIHVSFGI NL  LTT
Sbjct: 665 YQAEIDIYGLNRSGLLNDVLQILSNSTKSISTVNAQPTKDMKFANIHVSFGIPNLTHLTT 724

Query: 720 VVDKIKIIPDVYSVKRTNG 738
           VV+KIK +PDVYSVKRTNG
Sbjct: 725 VVEKIKAVPDVYSVKRTNG 743
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 501

A DNA sequence (GBSx0539) was identified in *S.agalactiae* <SEQ ID 1605> which encodes the amino acid sequence <SEQ ID 1606>. This protein is predicted to be 2',3'-cyclic-nucleotide 2'-phosphodiesterase precursor (cpdB). Analysis of this protein sequence reveals the following:

```
Possible site: 28
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -5.79   Transmembrane  779 - 795 ( 778 - 797)

----- Final Results -----
              bacterial membrane --- Certainty=0.3314(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB12613 GB:Z99108 similar to 2',3'-cyclic-nucleotide
           2'-phosphodiesterase [Bacillus subtilis]
 Identities = 297/630 (47%), Positives = 419/630 (66%), Gaps = 21/630 (3%)

Query: 102 KVDLRIMSTTDLHTNLVNYDYYQDKESQKIGLAKTAVLIEEAKKENPNTVLVDNGDVIQG 161
           +V L I++TTD+H N+++YDYY DKE+   GLA+TA LI++ +++NPNT+LVDNGD+IQG
```

```
Sbjct: 42  QVHLSILATTDIHANMMDYDYYSDKETADFGLARTAQLIQKHREQNPNTLLVDNGDLIQG 101

Query: 162 TPLGTYKAIVKP---VAENEEHPMYQAMNALGYDASTLGNHEFNYGLDYLKKIIATANLP 218
            PLG Y    +    ++  + HP+   MNAL YDA TLGNHEFNYGLD+L   I  A+ P
Sbjct: 102 NPLGEYAVKYQKDDIISGTKTHPIISVMNALKYDAGTLGNHEFNYGLDFLDGTIKGADFP 161

Query: 219 ILNANVLDFKTHQPVFKTYDIITKTFKDSTGRAVALNIGITGIVPPQILNWDKANLEGKV 278
           I+NANV    + +  +  Y I  KT  D  G    + +G  G VPPQI+ WDK NLEG+V
Sbjct: 162 IVNANVKT-TSGENRYTPYVINEKTLIDENGNEQKVKVGYIGFVPPQIMTWDKKNLEGQV 220

Query: 279 IVKDSVKAIEEIVPTMRAKGADVILVLSHSGIGDDRYEEGEENVGYQIAS-IKGVDAVVT 337
            V+D V++  E +P M+A+GADVI+ L+H+GI       G EN  + +A+  KG+DA+++
Sbjct: 221 QVQDIVESANETIPKMKAEGADVIIALAHTGIEKQAQSSGAENAVFDLATKTKGIDAIIS 280

Query: 338 GHSHAEFPSGNGTGFYEKYTGVDGIN---GKINGTPVTMAGKYGDHLGIIDLGLSYTNGK 394
           GH H  FPS        +Y GV   N   G ING PV M   +G +LG+IDL L   +G
Sbjct: 281 GHQHGLFPSA-------EYAGVAQFNVEKGTINGIPVVMPSSWGKYLGVIDLKLEKADGS 333

Query: 395 WQVSESSAKIRKIDMNSTTADERIIALAKEAHDGTINYVRQQVGTTTAPITSYFALVKDD 454
           W+V++S   I  I  N T+ +E +    ++ H  T+ YVR+ VG T A I S+FA VKDD
Sbjct: 334 WKVADSKGSIESIAGNVTSRNETVTNTIQQTHQNTLEYVRKPVGKTEADINSFFAQVKDD 393

Query: 455 PSVQIVNNAQRWYVANELKGTPEANLPLLSAAAPFKAGTRGDATAYTDIPAGPVAIKNVA 514
           PS+QIV +AQ+WY   E+K T   NLP+LSA APFKAG R  A  YT+IPAG +AIKNV
Sbjct: 394 PSIQIVTDAQKWYAEKEMKDTEYKNLPILSAGAPFKAGGRNGANYYTNIPAGDLAIKNVG 453

Query: 515 DLYLYDNVTALLKVTGADLREWLEMSAGQFNQIDPNNKAPQNIINTEYRTYNFDVIDGLT 574
           DLYLYDN   ++K+TG+++++WLEMSAGQFNQIDP     Q ++N  +R+YNFDVIDG+T
Sbjct: 454 DLYLYDNTVQIVKLTGSEVKDWLEMSAGQFNQIDPAKGGDQALLNENFRSYNFDVIDGVT 513

Query: 575 YKFDITQPNKYNKDGKVVNSQASRVRDLMYNGKPVADKQEFMIVTNNYRASGTFPGAKNA 634
           Y+ D+T+P KYN++GKV+N+ +SR+ +L Y GKP++  QEF++VTNNYRASG   G  +
Sbjct: 514 YQVDVTKPAKYNENGKVINADSSRIINLSYEGKPISPSQEFLVVTNNYRASGG-GGFPHL 572

Query: 635 TMNRLLN---LENRQTIINYIISEKTINPTADNNWGFTESIKDLDLRFQTADKAKNLVTN 691
           T+++++    +ENRQ +++YII +KT+NP ADNNW     +   +L F+++  AK
Sbjct: 573 TSDKIVHGSAVENRQVLMDYIIEQKTVNPKADNNWSIA-PVSGTNLTFESSLLAKPFADK 631

Query: 692 SKDIQYIASSTKDEGFGDYRFVYTEQEKVD 721
           + D+ Y+  S  +EG+G Y+  + +    D
Sbjct: 632 ADDVAYVGKSA-NEGYGVYKLQFDDDSNPD 660
 Identities = 133/567 (23%), Positives = 214/567 (37%), Gaps = 147/567 (25%)

Query: 104 DLRIMSTTDLHTNLVNYDYYQDKESQKIGLAKTAVLIEEAKKENPNTVLVDNGDVIQGTP 163
           DL +M T D H +L +              A+    I E + E  + +L+D GDV  G
Sbjct: 668 DLTVMHTNDTHAHLDD-------------AARRMTKINEVRSETNHNILLDAGDVFSGD- 713

Query: 164 LGTYKAIVKPVAENEEHPMYQAMNALGYDASTLGNHEFNYG----LDYLKKIIATAN--- 216
              Y      +A+       + MN +GYDA T GNHEF+ G     D+L    AT +
Sbjct: 714 --LYFTKWNGLAD------LKMMNMMGYDAMTFGNHEFDKGPTVLSDFLSGNSATVDPAN 765

Query: 217 --------LPILNANVLDFKTHQPVFKTYDIITKTF----KDSTGRAVALNIGITG--IV 262
                    PI++ANV    +++P  K++    +TF    K   G    + + + G  +
Sbjct: 766 RYHFEAPEFPIVSANV--DVSNEPKLKSFVKKPQTFTAGEKKEAGIHPYILLDVDGEKVA 823

Query: 263 PPQILNWDKANLE--GKVIV--------KDSVKAIEEIVPTMRAKGADVILVLSHSGIGD 312
              +   D A     GK IV        +++VKAI+E       +  + I+ L+H G
Sbjct: 824 VFGLTTEDTATTSSPGKSIVFNDAFETAQNTVKAIQE------EEKVNKIIALTHIG--- 874

Query: 313 DRYEEGEENVGYQIA-SIKGVDAVVTGHSHAEFPSGNGTGFYEKYTGVDGINGKINGTP- 370
                   N   ++A  +KG+D ++ GH+H               T VD +    N  P
Sbjct: 875 -------HNRDLELAKKVKGIDLIIGGHTH---------------TLVDKMEVVNNEEPT 912

Query: 371 -VTMAGKYGDHLGIIDLGLSYTNGKWQVSESSAKIRKIDMNSTTADERIIALAKEAHDGT 429
            V  A +YG  LG +D+      G  Q  +S+  +  ID ++    E     AK+  D
Sbjct: 913 IVAQAKEYGQFLGRVDVAFD-EKGVVQTDKSNLSVLPIDEHTEENPE-----AKQELDQF 966

Query: 430 INYV----RQQVGTTTAPITSYFALVKDDPSVQIVNNAQRWYVANELKGTPEANLPLLSA 485
```

```
             N +     ++VG T                  +  + QR +V  +           + A
Sbjct: 967  KNELEDVKNEKVGYT-----------------DVALDGQREHVRTKETNLGNFIADGMLA 1009

Query: 486  AAPFKAGTRGDAT----AYTDIPAGPVAIKNVADLYLYDNVTALLKVTGADLREWLEMSA 541
             A   AG R   T        I  G + +  V ++  + N   +  +TG  ++E LE
Sbjct: 1010 KAKEAAGARIAITNGGGIRAGIDKGDITLGEVLNVMPFGNTLYVADLTGKQIKEALE--- 1066

Query: 542  GQFNQIDPNNKAPQNIINTEYRTYNFDVIDGLTYKFDITQPNKYNKDGKVVNSQASRVRD 601
                          Q + N E     F  + G+ Y F +      NK G     +   V+
Sbjct: 1067 -------------QGLSNVENGGGAFPQVAGIEYTFTLN-----NKPG----HRVLEVKI 1104

Query: 602  LMYNGKPVADKQE--FMIVTNNYRASG 626
                NG  VA   +  + + TNN+  +G
Sbjct: 1105 ESPNGDKVAINTDDTYRVATNNFVGAG 1131
```

There is also homology to SEQ ID 1608. A related sequence was also identified in GAS <SEQ ID 9129> which encodes the amino acid sequence <SEQ ID 9130>. Analysis of this protein sequence reveals the following:

```
Possible cleavage site: 27

>>> Seems to have an uncleavable N-term signal seq
     INTEGRAL    Likelihood = -4.67   Transmembrane  649 - 665 ( 648 - 666)
     INTEGRAL    Likelihood = -2.02   Transmembrane    6 -  22 (   5 -  22)
     PERIPHERAL  Likelihood =  1.85
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

A related GBS gene <SEQ ID 8585> and protein <SEQ ID 8586> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1    Crend: 7
McG: Discrim Score:      6.68
GvH: Signal Score (-7.5): 0.87
     Possible site: 28
>>> Seems to have a cleavable N-term signal seq.
ALOM program   count: 1 value:  -5.79 threshold:  0.0
   INTEGRAL    Likelihood = -5.79   Transmembrane  779 - 795 ( 778 - 797)
   PERIPHERAL  Likelihood =  0.53     251
 modified ALOM score:   1.66

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.3314(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>

LPXTG motif: 769-773
```

The protein has homology with the following sequences in the databases:

```
ORF01378(298 - 2337 of 3000)
GP|6782402|emb|CAB70615.1||AJ133440(1  -  680  of  683)  cyclo-nucleotide  phosphodiesterase,
putative {Strept
ococcus dysgalactiae subsp. equisimilis}
%Match = 38.3
%Identity = 59.0  %Similarity = 72.3
Matches = 403  Mismatches = 181  Conservative Sub.s = 91

105       135       165       195       225       255       285       315
LFYHFLT*K*KKLEAQKELXTK*MCLTKLSFINKRLFLV*SLKIIRK*D*LNVFNKL**FL *DNIHVMF*WRRFMSKHY
                                                                           |:| |
                                                                          MMTKGY
```

```
345       375       405       435       465       495       525       555
FSKSVFALTVLTATATSGLAVQAEDIVTTPSSTSTKVESTTPTSTIAEEKSNVTSTPTAITDASTATSTNTTTNKQNPQP
 |||   | :| |  :: |  :||:       |:|   :|| |:  ||    |  |  |   | ||  :   : :
MSKSAIFLAMLVAAGSAQLT-KAEETTAVEPLTTT-ANTTTSTAVPAETAPLVADTTPATATADTAVPSPVNSTSSE-MA
            20         30        40         50        60        70        80

585       615       645       675       705       735       765       795
VATEATTSDLKPIEGEKVDLRIMSTTDLHTNLVNYDYYQDKESQKIGLAKTAVLIEEAKKENPNTVLVDNGDVIQGTPLG
 |:    :   |:||: ||:||:||||||:||||||||||||:| :|||||||||: ||||| | |||||||::|||||
TASADNATVTAPVEGQSVDVRILSTTDLHSNLVNYDYYQDKEAQSLGLAKTAVLIDAAKKENSNVVLVDNGDILQGTPLA
              100       110       120       130       140       150       160

825       855       885       915       945       975      1005      1035
TYKAIVKPVAENEEHPMYQAMNALGYDASTLGNHEFNYGLDYLKKIIATANLXILNANVLDFKTHQPVFKTYDIITKTFK
|||||| ||  :| |||| |: || :||||||||||||||||| :::||| | | | |||| || :| || :|||||||
TYKAIVDPVEADEVHPMYAALKALNFDASTLGNHEFNYGLDYLDRVMATAGLPIGNVNVLDAKTGKPKFKPFDIITKTFT
              180       190       200       210       220       230       240

1065      1095      1125      1155      1185      1215      1243      1273
DSTGRAVALNIGITGIVPPQILNWDKANLEGKVIVKDSVKAIEEIVPTMRAKGADVILVLS TLALEMIDMKKVKKTLAI
|  |: |:| |||||:|||||::|||||| ||| |||:|:|::|::||:|| |||::|||: ||   |  || ||| |
DKDGKTVSLKIGITGVVPPQIMSWDKANLTGKVTVKDAVEAVKEVIPTIRAAGADLVLVLA!TLVSVMTSMKSVKKMLVT
              260       270       280       290       300        310       320

1303                1357      1387      1416      1446      1476      1500
KLPASREWMPLLRDTHTL--NFHQVTVLASMKNTLELMVSM KINGTPVTMAGKYGDHLGIIDLGLSYTNGKWQVS--ES
|| | :  |   :  | |    :|: |:      | :  ::  ||| |||| ||||||||:||| |:||||:|:|:  :|
KLLALKVLMQWSQAIHMLISQPYQMAVFTITSKVLMVKRAL!-INGVPVTMGGKYGDHLGLIDLNLTYTNGQWKVNKDQS
               340       350       360         370       380       390       400

1530      1560      1590      1620      1650      1680      1710      1740
SAKIRKIDMNSTTADERIIALAKEAHDGTINYVRQQVGTTTAPITSYFALVKDDPSVQIVXNAQRWYVANELKGTPEANL
 |: |:||  |   |  ||||||||||||: ||||||||||||| |||||:|||||:||| ||||||   || |||||||
RAETRQIDSKSNQVDPTIIALAKEAHDGTVAYVRQQVGTTTAPINSYFALIKDDPSIQIVNNAQRWYAEKELAGTPEANL
       410       420       430       440       450       460       470       480

1770      1800      1830      1860      1887      1917      1947      1977
PLLSAAAPFKAGTRGDATAYTDIPAGPVAIKNVADLYLYDNVTAL-LKVTGADLREWLEMSAGQFNQIDPNNKAPQNIIN
||||||||||||          |     ::  ::  :     : | ||||||||:||||||||||| |||:   ||:::|
PLLSAAAPFKAGYTKMMRQLILIFLLVQSLSKMSLTFTCTTTSLLFLKVTGADLKEWLEMSAGQFNTIDPSKSEPQDLVN
       490       500       510       520       530       540       550       560

2007      2037      2067      2097      2127      2157      2187      2217
TEYRTYNFDVIDGLTYKFDITQPNKYNKDGKVVNSQASRVRDLMYNGKPVADKQEFMIVTNNYRASGTFPGAKNATMNRL
| ||||||||||||||:||:|| |||:  | :||  |||||:| | || :  |||||:||||||||| ||| ||||:|||
TSYRTYNFDVIDGLTYEFDVTQKNKYDSKGNLVNPDASRVRNLKYMGKDIDPKQEFMVVTNNYRASGNFPGVKNATLNRL
       570       580       590       600       610       620       630       640

2247      2277      2307      2337      2367      2397      2427      2457
LNLENRQTIINYIISEKTINPTADNNWGFTESIKDLDLRFQTADKAKNLVTNSKDIQYIASSTKDEGFGDYRFVYTEQEK
||||||| |||||::||||||:||||| | ::|| |:|||
LNLENRQAIINYIVAEKTINPSADNNWYFADTIKGLNLRFLKR
       650       660       670       680
```

SEQ ID 8586 (GBS53) was expressed in *E.coli* as a His-fusion product. The purified protein is shown in Figure 196, lane 9.

Example 502

A DNA sequence (GBSx0540) was identified in *S.agalactiae* <SEQ ID 1609> which encodes the amino acid sequence <SEQ ID 1610>. Analysis of this protein sequence reveals the following:

```
Possible site: 24
>>> Seems to have no N-terminal signal sequence
```

```
----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0296(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 503

A DNA sequence (GBSx0541) was identified in *S.agalactiae* <SEQ ID 1611> which encodes the amino acid sequence <SEQ ID 1612>. Analysis of this protein sequence reveals the following:

```
Possible site: 18
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1504(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10195> which encodes amino acid sequence <SEQ ID 10196> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB12860 GB:Z99109 similar to glucanase [Bacillus subtilis]
 Identities = 212/345 (61%), Positives = 268/345 (77%), Gaps = 1/345 (0%)

Query: 30  SMETTLNYIKTLTSIPSPTGFTQTIMTYIIKELEAFGYSPIRTNKGGVMVSLKGKNDTKH 89
            S+  T+  IK L SIPSPTG T  ++ YI   L+ +    +R +KGG++ +L G++ ++H
Sbjct: 3   SVRKTMELIKELVSIPSPTGNTYEVINYIESLLKEWKVETVRNHKGGLIATLPGRDTSRH 62

Query: 90  RMITAHLDTLGAMVRAIKPDGRLKIDLIGGYTYNAIEGENCTIHLSKNGKEISGTALIHQ 149
            RM+TAH+DTLGAMV+ IK DGRLKIDLIGG+ YN+IEGE C I  + +GK  +GT L+HQ
Sbjct: 63  RMLTAHVDTLGAMVKEIKADGRLKIDLIGGFRYNSIEGEYCQIETA-SGKMYTGTILMHQ 121

Query: 150 TSVHVYKDAGTAERNQTNMEIRLDEKVTTADETRALGIQVGDFISFDPRTIITDSGFIKS 209
            TSVHVYKDAG AERNQ NMEIRLDE V    +T  LGI VGDF+SFDPR  IT SGFIKS
Sbjct: 122 TSVHVYKDAGKAERNQENMEIRLDEPVHCRKDTEELGIGVGDFVSFDPRVEITSSGFIKS 181

Query: 210 RYLDDKVSAGILMELLSVYKKEDIQLPYTTHFYFSAFEELGHGANSSIPNETVEYLAVDM 269
            R+LDDK S  +L+ L+   + EDI+LPYTTHF  S  EE+G+G NS+IP ETVEYLAVDM
Sbjct: 182 RHLDDKASVALLLRLIHEIQTEDIELPYTTHFLISNNEEIGYGGNSNIPPETVEYLAVDM 241

Query: 270 GAMGDDQETDEYTVSICVKDASGPYHYELRQHLVSLAENNNIPYKLDIYPYYGSDASAAM 329
            GA+GD Q TDEY+VSICVKDASGPYHY+LR+HLV LAE ++I YKLDIYPYYGSDASAA+
Sbjct: 242 GAIGDGQATDEYSVSICVKDASGPYHYQLRKHLVQLAEKHHIDYKLDIYPYYGSDASAAI 301

Query: 330 RAGAEVKHALLGAGIESSHSYERTHIDSIQATELLVDAYLKSNMV 374
            ++G ++ H L+G GI++SH++ERTH  S++ T  L+  Y++S MV
Sbjct: 302 KSGHDIVHGLIGPGIDASHAFERTHKSSLRHTAKLLYYYVQSPMV 346
```

There is also homology to SEQ ID 424.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 504

A DNA sequence (GBSx0542) was identified in *S.agalactiae* <SEQ ID 1613> which encodes the amino acid sequence <SEQ ID 1614>. Analysis of this protein sequence reveals the following:

```
Possible site: 20
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3157(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAF11472 GB:AE002031 conserved hypothetical protein [Deinococcus radiodurans]
 Identities = 55/150 (36%), Positives = 85/150 (56%), Gaps = 2/150 (1%)

Query: 5   LIIIRGNSASGKSTIAKQLQAELGENTLLLSQDYLRREMLGTKDGENTTTIPLLINLLNY 64
            LI++RGNS SGKS++A+ L+   G     + QDYLRR +L   D      I L+   + Y
Sbjct: 23  LIVLRGNSGSGKSSVARALRERFGYGLAWVEQDYLRRVLLREHDVAGGKNIGLIETNVRY 82

Query: 65  GYHNCSYIILEGILRSDWYTPVWKHILKHNPNNTYAYYYDLSFQETVKRHSTRLKSLEFG 124
                 S  +LEGIL S  Y P+ + +  H     + +Y+DL F+ETV+RH+TR ++ +FG
Sbjct: 83  CLSAGSVTVLEGILFSRHYGPMLERL--HADFGGHWFYFDLPFEETVRRHATRPQAADFG 140

Query: 125 EDSLARWWLEKDFLKEIPEKILTKAMSLED 154
               +  W+  +D L  + E+++  A SL D
Sbjct: 141 VQDMQAWFQARDVLPFVQEQLIGPASSLAD 170
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 505

A DNA sequence (GBSx0543) was identified in *S.agalactiae* <SEQ ID 1615> which encodes the amino acid sequence <SEQ ID 1616>. This protein is predicted to be periplasmic-iron-binding protein BitC. Analysis of this protein sequence reveals the following:

```
Possible site: 29
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-11.46    Transmembrane    9 -  25 (   5 -  30)

----- Final Results -----
               bacterial membrane --- Certainty=0.5585(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAD18094 GB:U75349 periplasmic-iron-binding protein BitA
          [Brachyspira hyodysenteriae] (ver 2)
 Identities = 114/331 (34%), Positives = 184/331 (55%), Gaps = 3/331 (0%)

Query: 11  YILLVVSIIFISVFTYSISQPSKLLPPKELVILSPNSQAILTGTIPAFEEKY-GIKVKLI 69
            +I+  + ++ +++F    S  SK      LVI   +   ++   +  F+ K   I V+++
Sbjct: 4   FIIFCMLMLSMTLFYSCSSGDSK--NANSLVIYCSHPLDLMNTILDDFKAKNPDINVEVV 61

Query: 70  QGGTGQLIDRLSKEGKQLKADIFFGGNYTQFESHKALFESYVSKNVHTVIPDYIHPSDTA 129
              GTG+L+ R+  E      D+ +GG     +S   LFE+Y S N   ++ ++ +
Sbjct: 62  TAGTGELLKRVEAEKMNPLGDVLWGGTLNSVKSKTDLFENYTSTNEANILDEFKNTEGPF 121
```

```
Query: 130 TPYTINGSVLIVNNELAKGLTIKSYEDLLQPSLKGKIAFADPNTSSSAFSQLTNILLAKG 189
           T ++   S+L+VN  LA  + I+ YEDLL P LKGKIA ADP+ SSSAF  L N+L A G
Sbjct: 122 TRFSAIPSILMVNTNLAGNIKIEGYEDLLNPELKGKIAAADPSASSSAFEHLVNMLYAMG 181

Query: 190 GYTNPKAWNYVKKLQHNINAIKSSSSSEVYQSVAEGKMIVGLTYEDPSVNLQKSGANVSI 249
                K W+YV+KL  N++    S SS VY+ VA+G+  VGLTYE+P ++   SG+ V +
Sbjct: 182 KGDPEKGWDYVQKLCANLDGKLLSGSSAVYKGVADGEYTVGLTYEEPGISYMSSGSPVKV 241

Query: 250 VYPTEGTVFVPSSVAIIKNAPSMKEAKLFINFMLSLDVQNAFGQSTSNRPIRKDAQTSNG 309
           +Y  EG +  P  V IIK   +++ AK FI++ +SLD QN   +  S R IR DA  ++
Sbjct: 242 IYMKEGVISKPDGVYIIKGGKNLENAKKFIDYCVSLDAQNMLVEKLSRRSIRSDAVVTDM 301

Query: 310 MKALKDIATLKEDYRYVTKHKGQILKTYNRI 340
           +K + +I ++ ++   V + + + L  +  I
Sbjct: 302 VKPMSEIYSITDNADVVEESRQKWLDKFKDI 332
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1617> which encodes the amino acid sequence <SEQ ID 1618>. Analysis of this protein sequence reveals the following:

```
        Possible site: 33

>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-13.16    Transmembrane    9 -  25 (   4 -  33)

----- Final Results -----
                 bacterial membrane --- Certainty=0.6265(Affirmative) < succ>
                  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
                bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAB95371 GB:U75349 periplasmic-iron-binding protein BitC
           [Brachyspira hyodysenteriae]
 Identities = 115/324 (35%), Positives = 177/324 (54%), Gaps = 8/324 (2%)

Query: 15  VIIILAIVNVAMYIF-----SSSKKDSAKELVILTPNSQTILTGTIPAFEEKY-GVKVRL 68
           +++I   + ++++IF     S S   S   LVI  P+    +   +  F+ K  G+ V +
Sbjct: 4   IVLIFTSLLLSVFIFYSCSSSESGAQSGNSLVIYCPHPLEFINPLVDDFKAKNPGINVDI 63

Query: 69  IQGGTGQLIDQL-GRKDKPLNADIFFGGNYTQFESHKDLFESYVSPQVSTVISDYQLPSH 127
           I  GTG+L+ ++   KD PL  DI +GG  +  +   DLFESY S     +   Y+
Sbjct: 64  IAAGTGELLKRVESEKDNPLG-DILWGGTISMAKPKIDLFESYTSTNEENIAEIYKNTEG 122

Query: 128 RATPYTINGSVLIVNNELARGLHITSYEDLLQPALKGKIAFADPNSSSSAFSQLTNILLA 187
             T  T   S+L+VN  LA  + I  YEDLL P LKGKIAFADP++SSS+F  L N+L A
Sbjct: 123 ALTRCTAVPSILMVNTNLAGDIKIEGYEDLLNPELKGKIAFADPSASSSSFEHLVNMLYA 182

Query: 188 KGGYTNADAWAYMKRLLVNMNSIRATSSSEVYQSVAEGKMIVGLTYEDPCINLQKSGANV 247
            G       W Y+ +L  N++    + SS VY+ VA+G+  VGLT+E+   N   +G+ V
Sbjct: 183 IGKGDPEKGWDYVSKLCANLDGKLLSGSSAVYKGVADGEYTVGLTFEEGGANYVSAGSPV 242

Query: 248 SIVYPKEGTVFVPSSVAIIKHAPNMTEAKLFINFMLSRDVQNAFGQSTSNRPIRQDAQTS 307
            +VY KEG +  P  + IIK+A N+  AK F+++  S D Q       + R +R D   S
Sbjct: 243 KLVYMKEGVIIKPDGIYIIKNAKNLENAKKFVDYATSYDAQKTITDKLNRRSVRGDLPPS 302

Query: 308 HDMKALETIATLKEDYAYVTKHKK 331
             +++++TI  + +D A V ++K+
Sbjct: 303 AILQSVDTINVITDDEAVVDQNKQ 326
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 257/345 (74%), Positives = 295/345 (85%), Gaps = 1/345 (0%)

Query: 1   MKEKQSKRLIYILLVVSIIFISVFTYSISQPSKLLPPKELVILSPNSQAILTGTIPAFEE 60
           +K K+   L ++L+++ +  ++V  Y  S  SK    KELVIL+PNSQ ILTGTIPAFEE
Sbjct: 2   LKLKRKWLLSFLLVIIILAIVNVAMYIFSS-SKKDSAKELVILTPNSQTILTGTIPAFEE 60
```

```
Query: 61  KYGIKVKLIQGGTGQLIDRLSKEGKQLKADIFFGGNYTQFESHKALFESYVSKNVHTVIP 120
           KYG+KV+LIQGGTGQLID+L ++ K L ADIFFGGNYTQFESHK LFESYVS  V TVI
Sbjct: 61  KYGVKVRLIQGGTGQLIDQLGRKDKPLNADIFFGGNYTQFESHKDLFESYVSPQVSTVIS 120

Query: 121 DYIHPSDTATPYTINGSVLIVNNELAKGLTIKSYEDLLQPSLKGKIAFADPNTSSSAFSQ 180
           DY  PS  ATPYTINGSVLIVNNELA+GL I SYEDLLQP+LKGKIAFADPN+SSSAFSQ
Sbjct: 121 DYQLPSHRATPYTINGSVLIVNNELARGLHITSYEDLLQPALKGKIAFADPNSSSSAFSQ 180

Query: 181 LTNILLAKGGYTNPKAWNYVKKLQHNINAIKSSSSSEVYQSVAEGKMIVGLTYEDPSVNL 240
           LTNILLAKGGYTN  AW Y+K+L  N+N+I+++SSSEVYQSVAEGKMIVGLTYEDP +NL
Sbjct: 181 LTNILLAKGGYTNADAWAYMKRLLVNMNSIRATSSSEVYQSVAEGKMIVGLTYEDPCINL 240

Query: 241 QKSGANVSIVYPTEGTVFVPSSVAIIKNAPSMKEAKLFINFMLSLDVQNAFGQSTSNRPI 300
           QKSGANVSIVYP EGTVFVPSSVAIIK+AP+M EAKLFINFMLS DVQNAFGQSTSNRPI
Sbjct: 241 QKSGANVSIVYPKEGTVFVPSSVAIIKHAPNMTEAKLFINFMLSRDVQNAFGQSTSNRPI 300

Query: 301 RKDAQTSNGMKALKDIATLKEDYRYVTKHKGQILKTYNRIRRNAD 345
           R+DAQTS+ MKAL+ IATLKEDY YVTKHK +I+ TYN++R+  +
Sbjct: 301 RQDAQTSHDMKALETIATLKEDYAYVTKHKKKIVATYNQLRQRLE 345
```

SEQ ID 1616 (GBS263) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 49 (lane 4; MW 63kDa).

The GBS263-GST fusion product was purified (Figure 205, lane 5) and used to immunise mice. The resulting antiserum was used for FACS (Figure 301), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 506

A DNA sequence (GBSx0544) was identified in *S.agalactiae* <SEQ ID 1619> which encodes the amino acid sequence <SEQ ID 1620>. This protein is predicted to be response regulator. Analysis of this protein sequence reveals the following:

```
Possible site: 37
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4733(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAF31452 GB:AF221126 putative response regulator [Streptococcus pneumoniae]
 Identities = 85/252 (33%), Positives = 147/252 (57%), Gaps = 17/252 (6%)

Query: 2   YRLLIVEDEHLIRKWLRYAIDYQSLNILVVGEAKDGKEGAQLIQEEQPDIVLSDINMPIM 61
           Y +LIVEDE+L+R+ L   ++  + ++ ++G+A++G++  +LIQ++ PDI+L+DINMP +
Sbjct: 3   YTILIVEDEYLVRQGLTKLVNVAAYDMEIIGQAENGRQAWELIQKQVPDIILTDINMPHL 62

Query: 62  TAFDMFEATKGQSYAK---IILSGYADFPNAQSAIHYGVLEFLTKPLEKQALIDCLKTIM 118
               +    + ++Y +   + L+GY DF  A SA+  GV ++L KP  +Q + + L  I
Sbjct: 63  NGIQLASLVR-ETYPQVHLVFLTGYDDFDYALSAVKLGVDDYLLKPFSRQDIEEMLGKIK 121

Query: 119 ARIE-EHKEKHLQEHTELYLPLPQANDQVPEVIKDMLAWIHSHFHGKIVISQLAHDLGYS 177
            +++ E KE+ LQ+     L   +    + + I+  LA        +  +  LA DLG+S
Sbjct: 122 QKLDKEEKEEQLQD-----LLTNRFEGNMAQKIQSHLA------DSQFSLKSLASDLGFS 170

Query: 178 ESYLYTVTKKHLHITLSDYINQYRINQAIQLMFREPDLMVYQIAEAVGIYDYRYFDRVFK 237
            +YL ++ KK L +   DY+ + R+ QA +L+    DL +Y+IAE VG  D  YF + FK
```

```
Sbjct: 171 PTYLSSLIKKELGLPFQDYLVRERVKQA-KLLLLTTDLKIYEIAEKVGFEDMNYFTQRFK 229

Query: 238 KYLGQTVKAFKE 249
           +  G T + FK+
Sbjct: 230 QIAGVTPRQFKK 241
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1621> which encodes the amino acid sequence <SEQ ID 1622>. Analysis of this protein sequence reveals the following:

```
Possible site: 36
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4239(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 193/257 (75%), Positives = 226/257 (87%)

Query: 1   MYRLLIVEDEHLIRKWLRYAIDYQSLNILVVGEAKDGKEGAQLIQEEQPDIVLSDINMPI 60
           MY+L+I+EDEH+IRKWLRYAIDY++L+ILV+GEAKDGKEGA LI+E QPDIVL+DINMPI
Sbjct: 1   MYKLVIIEDEHIIRKWLRYAIDYKALDILVIGEAKDGKEGAVLIKESQPDIVLTDINMPI 60

Query: 61  MTAFDMFEATKGQSYAKIILSGYADFPNAQSAIHYGVLEFLTKPLEKQALIDCLKTIMAR 120
           MTAFDMFE TK Q+YAKIILSGYADFPNA+SAIHYGVLEFLTKP+EK AL +CL+TI+A+
Sbjct: 61  MTAFDMFEVTKDQTYAKIILSGYADFPNARSAIHYGVLEFLTKPIEKAALWECLQTIIAK 120

Query: 121 IEEHKEKHLQEHTELYLPLPQANDQVPEVIKDMLAWIHSHFHGKIVISQLAHDLGYSESY 180
           IE+ K  + +    +Y+PLPQ  DQ+PEV+KD+L W+H+HF  KI  S+LAHDLGYSESY
Sbjct: 121 IEKQKGSNQKTDACVYIPLPQMTDQIPEVVKDILEWVHAHFQDKISTSRLAHDLGYSESY 180

Query: 181 LYTVTKKHLHITLSDYINQYRINQAIQLMFREPDLMVYQIAEAVGIYDYRYFDRVFKKYL 240
           +Y   KKHL + LSDYINQYRINQAIQLM +EPDLMVY+IA+AVGIYDYRYFDRVFKKYL
Sbjct: 181 IYQNIKKHLQMPLSDYINQYRINQAIQLMQQEPDLMVYEIAQAVGIYDYRYFDRVFKKYL 240

Query: 241 GQTVKAFKEEHIFKQMD 257
           GQTVKAFKEEH  K  D
Sbjct: 241 GQTVKAFKEEHFMKDTD 257
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 507

A DNA sequence (GBSx0545) was identified in *S.agalactiae* <SEQ ID 1623> which encodes the amino acid sequence <SEQ ID 1624>. Analysis of this protein sequence reveals the following:

```
Possible site: 14
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2964(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 508

A DNA sequence (GBSx0546) was identified in *S.agalactiae* <SEQ ID 1625> which encodes the amino acid sequence <SEQ ID 1626>. This protein is predicted to be two-component sensor histidine kinase. Analysis of this protein sequence reveals the following:

```
Possible site: 45
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-13.80   Transmembrane  266 - 282 ( 257 - 285)
   INTEGRAL    Likelihood =-12.90   Transmembrane   29 -  45 (  24 -  51)

----- Final Results -----
               bacterial membrane --- Certainty=0.6519(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10197> which encodes amino acid sequence <SEQ ID 10198> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAB05628 GB:AP001513 two-component sensor histidine kinase
           [Bacillus halodurans]
 Identities = 84/258 (32%), Positives = 138/258 (52%), Gaps = 23/258 (8%)

Query: 298 SSAINQMVLDMDAISRQEKSSIELDSQDEFQYLSVQINQMVSRLKDLHEKTLDLETQKLL 357
           S  INQ+       S   K+ I +D +DE   LSVQ NQMV+ L+ L  +  +   QK L
Sbjct: 327 SERINQVA------SGDLKTKIVVDGKDEIGQLSVQFNQMVANLRSLIHQVHETNRQKRL 380

Query: 358 FEK-------RMLEAQFNPHFLYNTLETILITSHYDSQL-TERIVIQLTKLLRYSLSGST 409
            EK       +ML +Q NPHFL+NTLE+I + SH   +    ++V QL KL+R SL  +
Sbjct: 381 LEKSQNEIKLKMLASQINPHFLFNTLESIRMKSHMKGETEIAKVVKQLGKLMRKSLEVTG 440

Query: 410 EAAVLKDDLAIIESYLLINQVRF-EELTYTISVSPELEHMRVPKLFLLPLIENAIKYGLK 468
               L+++L ++  YL I   R+ + L Y + + P+ E + +  L + PL+ENA+ +GL+
Sbjct: 441 HHIPLRNELDMVRCYLEIQTFRYGDRLHYELYIDPQSEMVEILPLIIQPLVENAVIHGLE 500

Query: 469 ERHD-VAINIDIWQDSDGIWFTVSNNGSGISLARQQAIRTMLRSTH----SHHGLINSYR 523
              D   + I    + + +   V+++G G+   + +AI+ ML        +  GL+N ++
Sbjct: 501 RTEDGGTVTISTIVNGNDLTVIVNDDGCGMDEEKLEAIQNMLHHPQEVDGNKIGLLNVHK 560

Query: 524 RLQYQF---STVLLEFTK 538
           RLQ  +   S +++E  K
Sbjct: 561 RLQLTYGKTSGLIIESAK 578
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1627> which encodes the amino acid sequence <SEQ ID 1628>. Analysis of this protein sequence reveals the following:

```
     Possible site: 43

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-10.88   Transmembrane   27 -  43 (  22 -  49)
   INTEGRAL    Likelihood = -9.08   Transmembrane  263 - 279 ( 258 - 282)

----- Final Results -----
               bacterial membrane --- Certainty=0.5352(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:BAB05628 GB:AP001513 two-component sensor histidine kinase
           [Bacillus halodurans]
 Identities = 85/270 (31%), Positives = 139/270 (51%), Gaps = 20/270 (7%)
```

```
Query: 276 IFVILQRKSSGLANRIAAKNSRAINQMVRDMSAISRQEKRRIDLESQDEFQYLSDQINQM 335
           + V+L    S L ++  +  S  INQ+       S   K +I ++ +DE   LS Q NQM
Sbjct: 307 VAVLLIVHFSWLISKRLSHLSERINQVA------SGDLKTKIVVDGKDEIGQLSVQFNQM 360

Query: 336 VERLQQLHDKTLDLETQKLLFEK-------RMLEAQFNPHFLYNTLETILITSHYDSAL- 387
           V  L+ L  +  +   QK L EK       +ML +Q NPHFL+NTLE+I + SH
Sbjct: 361 VANLRSLIHQVHETNRQKRLLEKSQNEIKLKMLASQINPHFLFNTLESIRMKSHMKGETE 420

Query: 388 TEKIVIQLTKLLRYSLTDSSKPVLLKDDLSVIESYLVINQVRF-EELQYSINLSPDLDSL 446
             K+V QL KL+R SL  +   + L+++L ++  YL I   R+ + L Y + + P  + +
Sbjct: 421 IAKVVKQLGKLMRKSLEVTGHHIPLRNELDMVRCYLEIQTFRYGDRLHYELYIDPQSEMV 480

Query: 447 EVPKLFLLPLIENAIKYGLKERHD-VKINIACYYQDDHIIFSVRDNGSGIDAHHQKVIRE 505
           E+  L + PL+ENA+ +GL+   D   + I+   + +     V D+G G+D    + I+
Sbjct: 481 EILPLIIQPLVENAVIHGLERTEDGGTVTISTIVNGNDLTVIVNDDGCGMDEEKLEAIQN 540

Query: 506 QL----EAGESHHGLINSYRRLKYHFSEVS 531
            L    E   +  GL+N ++RL+  + + S
Sbjct: 541 MLHHPQEVDGNKIGLLNVHKRLQLTYGKTS 570
```

An alignment of the GAS and GBS proteins is shown below:

```
Identities = 369/549 (67%), Positives = 449/549 (81%)

Query: 3   MRGYRMEERFKKRLQDDISKHFSRQSLILSLLLIALFVLFSLAPQQIGLYKDVNSVSYSY 62
           MRG ++EE FKK+LQDDIS+HFS QSL+LSLLLI LF++FSLAPQQ+GLY+D+N+ +  Y
Sbjct: 1   MRGEQVEEHFKKQLQDDISRHFSYQSLMLSLLLIGLFIIFSLAPQQLGLYRDINATATRY 60

Query: 63  KQLIQKHDTLLDDLGKNSLKPFVSGHLGSADLSKQYYHLRNHLQSQTELLVFSPNQELLF 122
            +LI K + LLDDLGKNSL PF++ +L +ADLSK Y+HLR+  Q+  ELL+FSP+Q+LLF
Sbjct: 61  HRLISKQEALLDDLGKNSLLPFLNKNLSTADLSKHYFHLRHSSQTSPELLLFSPSQDLLF 120

Query: 123 ASNSHLGNFFSKSIYISEVLDKAKINQRLLKIIVDSEGGHYLALIKPIIVNKKVSGYAFL 182
           ASN HLGN FSKS+YI EVL      + L K  +DSE GHYL +I P+I   ++ GYAFL
Sbjct: 121 ASNPHLGNVFSKSVYIQEVLRATHSPKTLFKDAMDSEDGHYLMIIMPMIDQNQLKGYAFL 180

Query: 183 LMNGKDFLLPTKAINSDLIIADQLNNSFTFTNRDFISSSLDKVDSQFLTRYFSFHDHRAF 242
           +M+GKDFL PTK + S+L+IAD+L+N+FTF+NR+FI+SSLDK++SQ+L  YF F D+RAF
Sbjct: 181 VMSGKDFLHPTKTLTSELVIADKLDNTFTFSNREFIASSLDKINSQYLHHYFVFQDNRAF 240

Query: 243 VVRKVALQDNILLYMYRPLIPVTLVVLFSLVSSVIIFVILRQKSRVLADRIAVKNSSAIN 302
           + RKVALQ  + LYMYRPLIP+  V+LFSL+SS +IFVIL++KS  LA+RIA KNS AIN
Sbjct: 241 ITRKVALQGGLWLYMYRPLIPMVSVMLFSLISSAVIFVILQRKSSGLANRIAAKNSRAIN 300

Query: 303 QMVLDMDAISRQEKSSIELDSQDEFQYLSVQINQMVSRLKDLHEKTLDLETQKLLFEKRM 362
           QMV DM AISRQEK  I+L+SQDEFQYLS QINQMV RL+ LH+KTLDLETQKLLFEKRM
Sbjct: 301 QMVRDMSAISRQEKRRIDLESQDEFQYLSDQINQMVERLQQLHDKTLDLETQKLLFEKRM 360

Query: 363 LEAQFNPHFLYNTLETILITSHYDSQLTERIVIQLTKLLRYSLSGSTEAAVLKDDLAIIE 422
           LEAQFNPHFLYNTLETILITSHYDS LTE+IVIQLTKLLRYSL+ S++  +LKDDL++IE
Sbjct: 361 LEAQFNPHFLYNTLETILITSHYDSALTEKIVIQLTKLLRYSLTDSSKPVLLKDDLSVIE 420

Query: 423 SYLLINQVRFEELTYTISVSPELEHMRVPKLFLLPLIENAIKYGLKERHDVAINIDIWQD 482
           SYL+INQVRFEEL Y+I++SP+L+ + VPKLFLLPLIENAIKYGLKERHDV INI  +
Sbjct: 421 SYLVINQVRFEELQYSINLSPDLDSLEVPKLFLLPLIENAIKYGLKERHDVKINIACYYQ 480

Query: 483 SDGIWFTVSNNGSGISLARQQAIRTMLRSTHSHHGLINSYRRLQYQFSTVLLEFTKTDDA 542
            D I F+V +NGSGI    Q+ IR  L +  SHHGLINSYRRL+Y FS V L F + D
Sbjct: 481 DDHIIFSVRDNGSGIDAHHQKVIREQLEAGESHHGLINSYRRLKYHFSEVSLVFDQGDKQ 540

Query: 543 FRVSYIVKE 551
           F VSY VKE
Sbjct: 541 FNVSYHVKE 549
```

A related GBS gene <SEQ ID 8587> and protein <SEQ ID 8588> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 10
McG: Discrim Score:      6.23
GvH: Signal Score (-7.5): -0.0500002
     Possible site: 38
>>> Seems to have a cleavable N-term signal seq.
ALOM program   count: 1 value: -13.80 threshold:  0.0
   INTEGRAL    Likelihood =-13.80   Transmembrane  259 - 275 ( 250 - 278)
   PERIPHERAL  Likelihood =  2.70     404
 modified ALOM score:   3.26

*** Reasoning Step: 3

----- Final Results -----
              bacterial membrane --- Certainty=0.6519(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
33.2/53.9% over 181aa
                           Streptococcus pneumoniae
  GP|5830535| histidine kinase Insert characterized

ORF00032(1309 - 1848 of 2253)
GP|5830535|emb|CAB54576.1||AJ006396(1  -  182  of  231)  histidine  kinase  {Streptococcus
pneumoniae}
%Match = 5.9
%Identity = 33.2  %Similarity = 53.8
Matches = 61  Mismatches = 78  Conservative Sub.s = 38

1053      1083      1113      1143      1173      1203      1233      1263
FVVRKVALQDNILLYMYRPLIPVTLVVLFSLVSSVIIFVILRQKSRVLADRIAVKNSSAINQMVLDMDAISRQEKSSIEL

1293      1323      1350      1380      1410      1440                1494
DSQDEFQYLSVQINQMVSRL-KDLHEKTLDLETQKLLFEKRMLEAQFNPHFLYNTLETILITSHYDSQ--LTERIVIQLT
               |: || |::|:    ||  :     | |:|| ||||:||||| : : :   ||  | : |: :::
               MLDRLEKNIHD-IYQLELSQKDANMRALQAQINPHFMYNTLEFLRMYAVMQSQDELAD-IIYEFS
                       10         20        30        40        50         60

1524      1554      1584      1611      1641      1671      1701      1728
KLLRYSLSGSTEAAVLKDDLAIIESYLLINQVRF-EELTYTISVSPELEHMRVPKLFLLPLIENAIKYGLKERH-DVAIN
 ||| ::|      :|| :|     |  :  ||: : : |   : ||||:|::||: | ||:||   :|:  |  |  |:
SLLRNNIS-DERETLLKQELEFCRKYSYLCMVRYPKSIAYGFKIDPELENMKIPKFTLQPLVENYFAHGVDHRRTDNVIS
                80        90       100       110       120       130       140

1758      1788      1818      1848      1878      1908      1938      1968
IDIWQDSDGIWFTVSNNGSGISLARQQAIRTMLRSTHSHHGLINSYRRLQYQFSTVLLEFTKTDDAFRVSYIVKE*VMYR
|   :    :   | :|| |:|  :   ||  |   :  |    | :|       |   |           |
IKALKQDGFVEILVVDNGRGMSAEKLANIREKLSQRYFEHQASYSDQRQSIGIVNVHERFVLYFGDRYAITIESAEQAGV
               160       170       180       190       200       210       220
```

SEQ ID 8588 (GBS47) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 14 (lane 2; MW 84kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 85 (lane 4; MW 59.3kDa).

GBS47-His was purified as shown in Figure 221, lane 4-5.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 509

A DNA sequence (GBSx0547) was identified in *S.agalactiae* <SEQ ID 1629> which encodes the amino acid sequence <SEQ ID 1630>. This protein is predicted to be phosphotransferase enzyme II, D component. Analysis of this protein sequence reveals the following:

```
Possible site: 32
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-10.46   Transmembrane  258 - 274 ( 252 - 274)
   INTEGRAL    Likelihood = -9.13   Transmembrane  232 - 248 ( 227 - 251)
   INTEGRAL    Likelihood = -5.31   Transmembrane  142 - 158 ( 140 - 161)
   INTEGRAL    Likelihood = -2.50   Transmembrane  119 - 135 ( 118 - 139)

----- Final Results -----
               bacterial membrane --- Certainty=0.5182(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAC74889 GB:AE000276 PTS enzyme IID, mannose-specific
           [Escherichia coli K12]
 Identities = 94/280 (33%), Positives = 156/280 (55%), Gaps = 13/280 (4%)

Query: 3   SQDNLTKEDRKMLRSVFWRSWTMNASRTGATQYHAVGVIYTLLPVINRFYKTDKD-KAEA 61
           ++  LT+ D   +R VF RS     S     +  A+G  ++++P I R Y  + + + +A
Sbjct: 12  TEKKLTQSD---IRGVFLRSNLFQGS-WNFERMQALGFCFSMVPAIRRLYPENNEARKQA 67

Query: 62  LVRHTTWFNATMHINNFIMGLVASMEKKNSEDPDFDASAITAVKASLMGPISGVGDSFFW 121
           + RH  +FN    +   I+G+  ++E++ +   + D  AI  +K  LMGP++GVGD  FW
Sbjct: 68  IRRHLEFFNTQPFVAAPILGVTLALEEQRANGAEIDDGAINGIKVGLMGPLAGVGDPIFW 127

Query: 122 GILRVIAAGIGISLASTGSAMGAVVFLLLYNIPAFLIHYYSLYGGYSVGAGFIKKLYESG 181
           G +R + A +G  +A +GS +G ++F +L+N+      YY +  GYS G   +K +   G
Sbjct: 128 GTVRPVFAALGAGIAMSGSLLGPLLFFILFNLVRLATRYYGVAYGYSKGIDIVKDM-GGG 186

Query: 182 GIKIVTKTSSMLGLMMVGSM----TASNVKFKTILTVAAKGAKEAASIQSYLDQLFVGVV 237
            ++ +T+ +S+LGL ++G++    T  N+           G +   ++Q+ LDQL  G+V
Sbjct: 187 FLQKLTEGASILGLFVMGALVNKWTHVNIPLVVSRITDQTGKEHVTTVQTILDQLMPGLV 246

Query: 238 PLLVTILAFWLLRKKVNINWIMFGIMVLGI---VLGLLGI 274
           PLL+T    WLLRKKVN  WI+ G  V+GI     GLLG+
Sbjct: 247 PLLLTFACMWLLRKKVNPLWIIVGFFVIGIAGYACGLLGL 286
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1631> which encodes the amino acid sequence <SEQ ID 1632>. Analysis of this protein sequence reveals the following:

```
     Possible site: 32
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -9.98   Transmembrane  255 - 271 ( 251 - 274)
   INTEGRAL    Likelihood = -7.01   Transmembrane  232 - 248 ( 228 - 250)
   INTEGRAL    Likelihood = -5.68   Transmembrane  142 - 158 ( 140 - 161)
   INTEGRAL    Likelihood = -2.50   Transmembrane  119 - 135 ( 118 - 139)

----- Final Results -----
               bacterial membrane --- Certainty=0.4991(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAC74889 GB:AE000276 PTS enzyme IID, mannose-specific
           [Escherichia coli]
 Identities = 94/281 (33%), Positives = 157/281 (55%), Gaps = 13/281 (4%)

Query: 2   TSQDNLTKEDRKMLRSVFWRSWTMNASRTGATQYHAVGVIYTLLPVINRFYKTDKD-KAE 60
```

```
           T++  LT+ D   +R VF RS     S     +  A+G  ++++P I R Y  + + + +
Sbjct: 11  TTEKKLTQSD---IRGVFLRSNLFQGS-WNFERMQALGFCFSMVPAIRRLYPENNEARKQ 66

Query: 61  ALVRHTTWFNATMHINNFIMGLVASMEKKNSEDPDFDASAITAVKASLMGPISGVGDSFF 120
           A+ RH  +FN    +   I+G+  ++E++ +   + D  AI  +K  LMGP++GVGD  F
Sbjct: 67  AIRRHLEFFNTQPFVAAPILGVTLALEEQRANGAEIDDGAINGIKVGLMGPLAGVGDPIF 126

Query: 121 WGILRVIAAGIGISLASAGSAMGAVVFLLLYNIPAFIIHYYSLYGGYSVGAGFIKKLYES 180
           WG +R + A +G  +A +GS +G ++F +L+N+      YY +  GYS G   +K +
Sbjct: 127 WGTVRPVFAALGAGIAMSGSLLGPLLFFILFNLVRLATRYYGVAYGYSKGIDIVKDM-GG 185

Query: 181 GGIKIVTKTSSMLGLMMVGSM----TASNVKFKTILTVAAKGAKEAASIQDYLDQLFIGI 236
           G ++ +T+ +S+LGL ++G++    T  N+           G +   ++Q  LDQL  G+
Sbjct: 186 GFLQKLTEGASILGLFVMGALVNKWTHVNIPLVVSRITDQTGKEHVTTVQTILDQLMPGL 245

Query: 237 VPLMVTLAAFWLLRKKVNIIWIMFGIMFLGI---ILGLLGI 274
           VPL++T A  WLLRKKVN +WI+ G   +GI     GLLG+
Sbjct: 246 VPLLLTFACMWLLRKKVNPLWIIVGFFVIGIAGYACGLLGL 286
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 263/275 (95%), Positives = 269/275 (97%)

Query: 1   MKSQDNLTKEDRKMLRSVFWRSWTMNASRTGATQYHAVGVIYTLLPVINRFYKTDKDKAE 60
           M SQDNLTKEDRKMLRSVFWRSWTMNASRTGATQYHAVGVIYTLLPVINRFYKTDKDKAE
Sbjct: 1   MTSQDNLTKEDRKMLRSVFWRSWTMNASRTGATQYHAVGVIYTLLPVINRFYKTDKDKAE 60

Query: 61  ALVRHTTWFNATMHINNFIMGLVASMEKKNSEDPDFDASAITAVKASLMGPISGVGDSFF 120
           ALVRHTTWFNATMHINNFIMGLVASMEKKNSEDPDFDASAITAVKASLMGPISGVGDSFF
Sbjct: 61  ALVRHTTWFNATMHINNFIMGLVASMEKKNSEDPDFDASAITAVKASLMGPISGVGDSFF 120

Query: 121 WGILRVIAAGIGISLASTGSAMGAVVFLLLYNIPAFLIHYYSLYGGYSVGAGFIKKLYES 180
           WGILRVIAAGIGISLAS GSAMGAVVFLLLYNIPAF+IHYYSLYGGYSVGAGFIKKLYES
Sbjct: 121 WGILRVIAAGIGISLASAGSAMGAVVFLLLYNIPAFIIHYYSLYGGYSVGAGFIKKLYES 180

Query: 181 GGIKIVTKTSSMLGLMMVGSMTASNVKFKTILTVAAKGAKEAASIQSYLDQLFVGVVPLL 240
           GGIKIVTKTSSMLGLMMVGSMTASNVKFKTILTVAAKGAKEAASIQ YLDQLF+G+VPL+
Sbjct: 181 GGIKIVTKTSSMLGLMMVGSMTASNVKFKTILTVAAKGAKEAASIQDYLDQLFIGIVPLM 240

Query: 241 VTILAFWLLRKKVNINWIMFGIMVLGIVLGLLGIC 275
           VT+ AFWLLRKKVNI WIMFGIM LGI+LGLLGIC
Sbjct: 241 VTLAAFWLLRKKVNIIWIMFGIMFLGIILGLLGIC 275
```

There is also homology to SEQ ID 5236.

A further related DNA sequence was identified in *S.pyogenes* <SEQ ID 9077> which encodes the amino acid sequence <SEQ ID 9078>. An alignment of the GAS and GBS sequences follows:

```
 Score =  178 bits (448), Expect = 3e-47
 Identities = 83/136 (61%), Positives = 108/136 (79%)

Query: 2   IMEEITIYHNPNCGTSRNVLAMIRHAGIEPTIIEYLQTPPNRETLIELLQSMGISARELL 61
           +ME+I IYHNPNCGTSRNVLA+IRH GIEP II YL+TPP+R  L+ELL  M +SARELL
Sbjct: 1   MMEKIRIYHNPNCGTSRNVLAIIRHCGIEPEIIYYLKTPPSRMELVELLLEMKLSARELL 60

Query: 62  RTNVPEFEAYGLANQAVAEKDIINAMLADPILINRPIVVTRKGVKLCRPSETLLDILPVP 121
           RT+VP +E + L + +V ++++I+AM+ DPILINRPIVVT KG KLCRP E +L ILPV
Sbjct: 61  RTDVPAYEKFNLESSSVTDEEMIDAMIQDPILINRPIVVTSKGAKLCRPCEAILTILPVK 120

Query: 122 LPSPYIKEDGESVNPI 137
           +   ++KEDG+ +  +
Sbjct: 121 MEKDFVKEDGQIIQSL 136
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 510

A DNA sequence (GBSx0548) was identified in *S.agalactiae* <SEQ ID 1633> which encodes the amino acid sequence <SEQ ID 1634>. This protein is predicted to be PTS permease for mannose subunit IIPMan. Analysis of this protein sequence reveals the following:

```
Possible site: 56
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -8.70   Transmembrane  144 - 160 ( 140 - 165)
   INTEGRAL    Likelihood = -8.07   Transmembrane  220 - 236 ( 215 - 239)
   INTEGRAL    Likelihood = -7.27   Transmembrane   95 - 111 (  91 - 116)
   INTEGRAL    Likelihood = -3.77   Transmembrane    2 -  18 (   1 -  18)
   INTEGRAL    Likelihood = -1.44   Transmembrane  180 - 196 ( 179 - 196)
   INTEGRAL    Likelihood = -1.33   Transmembrane   32 -  48 (  30 -  49)
   INTEGRAL    Likelihood = -0.53   Transmembrane  198 - 214 ( 198 - 214)

----- Final Results -----
               bacterial membrane --- Certainty=0.4482(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAC44680 GB:U65015 PTS permease for mannose subunit IIPMan
           [Vibrio furnissii]
 Identities = 70/251 (27%), Positives = 132/251 (51%), Gaps = 6/251 (2%)

Query: 2   IMPATMAALAVLICFGGNYLTGQSMMERPLVVGLVTGMLLGDIKVGILMGASLEALFLGN 61
           +  A M  L   +  G +   G +   RP+V+G + G++LGD+  GIL+G +LE +++G
Sbjct: 5   LFQALMLGLLAFLA-GLDLFNGLTHFHRPVVLGPLVGLILGDLHTGILVGGTLELIWMGL 63

Query: 62  VNIGGVIAAEPVTATAMATTFTIISNIDQKAAMTLAVPIGMLAAFVVMFLKNVFMNIFAP 121
             + G      +  T + TTF I +N++   A+ +AVP  +     +  L +    + +
Sbjct: 64  APLAGAQPPNVIIGTIVGTTFAITTNVEPNVAVGVAVPFAVAVQMGITLLFSAMSAVMSK 123

Query: 122 MVDKAAAANHQGKLVMLHYGTWII--YYLIIASISFIGILVGSGPVNSFVHHIPQNLMNG 179
             + A  A+ +G   + ++   ++  +Y + A   F+ I +G+    + V  +P+ L++G
Sbjct: 124 CDEYAKNADTRGIERVNYFALAVLGSFYFLCA---FLPIYLGADHAGAMVAALPKALIDG 180

Query: 180 LSAAGGLLPAVGFAMLMKLLWTNKLAVFYLLGFVLTAYLKLPAVAVAALGAVICVISSQR 239
           L  AGG++PA+GFA+LMK++  N    +++LGFV  A+L+LP +A+      + +I   R
Sbjct: 181 LGVAGGIMPAIGFAVLMKIMMKNAYIPYFILGFVAAAWLQLPILAIRCAATAMAIIDFMR 240

Query: 240 DIELDAITRGA 250
             E   +   A
Sbjct: 241 KSEPTPVNASA 251
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1635> which encodes the amino acid sequence <SEQ ID 1636>. Analysis of this protein sequence reveals the following:

```
     Possible site: 56
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -8.70   Transmembrane  144 - 160 ( 140 - 165)
   INTEGRAL    Likelihood = -8.07   Transmembrane  220 - 236 ( 215 - 239)
   INTEGRAL    Likelihood = -7.27   Transmembrane   95 - 111 (  91 - 116)
   INTEGRAL    Likelihood = -4.62   Transmembrane    2 -  18 (   1 -  19)
   INTEGRAL    Likelihood = -1.44   Transmembrane  180 - 196 ( 179 - 196)
   INTEGRAL    Likelihood = -0.96   Transmembrane   32 -  48 (  31 -  49)
   INTEGRAL    Likelihood = -0.53   Transmembrane  198 - 214 ( 198 - 214)

----- Final Results -----
               bacterial membrane --- Certainty=0.4482(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAC44680 GB:U65015 PTS permease for mannose subunit IIPMan
           [Vibrio furnissii]
 Identities = 72/251 (28%), Positives = 132/251 (51%), Gaps = 6/251 (2%)

Query: 2   LVPATMAALAVLICFGGNYLTGQSMMERPLVVGLVTGLLLGDMKVGILMGASLEALFLGN 61
           L  A M  L   +  G +   G +   RP+V+G + GL+LGD+  GIL+G +LE +++G
Sbjct: 5   LFQALMLGLLAFLA-GLDLFNGLTHFHRPVVLGPLVGLILGDLHTGILVGGTLELIWMGL 63

Query: 62  VNIGGVIAAEPVTATAMATTFTIISHIDQKAAMTLAVPIGMLAAFVVMFLKNVFMNIFAP 121
             + G      +  T + TTF I ++++   A+ +AVP  +     +  L +    + +
Sbjct: 64  APLAGAQPPNVIIGTIVGTTFAITTNVEPNVAVGVAVPFAVAVQMGITLLFSAMSAVMSK 123

Query: 122 MVDKAAAANHQGKLVMLHYGTWII--YYLIIASISFIGILVGSGPVNAFVEHIPQNLMNG 179
             + A  A+ +G   + ++   ++  +Y + A   F+ I +G+    A V  +P+ L++G
Sbjct: 124 CDEYAKNADTRGIERVNYFALAVLGSFYFLCA---FLPIYLGADHAGAMVAALPKALIDG 180

Query: 180 LSAAGGLLPAVGFAMLMKLLWTNKLAVFYLLGFVLTAYLKLPAVAVAALGAVICVISSQR 239
           L  AGG++PA+GFA+LMK++  N    +++LGFV  A+L+LP +A+      + +I   R
Sbjct: 181 LGVAGGIMPAIGFAVLMKIMMKNAYIPYFILGFVAAAWLQLPILAIRCAATAMAIIDFMR 240

Query: 240 DLELDAITRGA 250
             E   +   A
Sbjct: 241 KSEPTPVNASA 251
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 261/269 (97%), Positives = 268/269 (99%)

Query: 1   MIMPATMAALAVLICFGGNYLTGQSMMERPLVVGLVTGMLLGDIKVGILMGASLEALFLG 60
           M++PATMAALAVLICFGGNYLTGQSMMERPLVVGLVTG+LLGD+KVGILMGASLEALFLG
Sbjct: 1   MLVPATMAALAVLICFGGNYLTGQSMMERPLVVGLVTGLLLGDMKVGILMGASLEALFLG 60

Query: 61  NVNIGGVIAAEPVTATAMATTFTIISNIDQKAAMTLAVPIGMLAAFVVMFLKNVFMNIFA 120
           NVNIGGVIAAEPVTATAMATTFTIIS+IDQKAAMTLAVPIGMLAAFVVMFLKNVFMNIFA
Sbjct: 61  NVNIGGVIAAEPVTATAMATTFTIISHIDQKAAMTLAVPIGMLAAFVVMFLKNVFMNIFA 120

Query: 121 PMVDKAAAANHQGKLVMLHYGTWIIYYLIIASISFIGILVGSGPVNSFVHHIPQNLMNGL 180
           PMVDKAAAANHQGKLVMLHYGTWIIYYLIIASISFIGILVGSGPVN+FV HIPQNLMNGL
Sbjct: 121 PMVDKAAAANHQGKLVMLHYGTWIIYYLIIASISFIGILVGSGPVNAFVEHIPQNLMNGL 180

Query: 181 SAAGGLLPAVGFAMLMKLLWTNKLAVFYLLGFVLTAYLKLPAVAVAALGAVICVISSQRD 240
           SAAGGLLPAVGFAMLMKLLWTNKLAVFYLLGFVLTAYLKLPAVAVAALGAVICVISSQRD
Sbjct: 181 SAAGGLLPAVGFAMLMKLLWTNKLAVFYLLGFVLTAYLKLPAVAVAALGAVICVISSQRD 240

Query: 241 IELDAITRGAISKQTTFDSKESEEEDFFA 269
           +ELDAITRGAISKQTTFDSKESEEEDFFA
Sbjct: 241 LELDAITRGAISKQTTFDSKESEEEDFFA 269
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 511

A DNA sequence (GBSx0549) was identified in *S.agalactiae* <SEQ ID 1637> which encodes the amino acid sequence <SEQ ID 1638>. This protein is predicted to be pts system, sorbose-specific iib component. Analysis of this protein sequence reveals the following:

```
Possible site: 36
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1874(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
```

```
                  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAA46858 GB:X66059 EIII-B Sor PTS [Klebsiella pneumoniae]
 Identities = 49/158 (31%), Positives = 94/158 (59%), Gaps = 8/158 (5%)

Query: 2   ITQIRVDDRLIHGQVAVVWTKELNAPLLVVANDEAAKNEITQMTLKMAVPNGMKLLIRSV 61
            IT  R+DDRLIHGQV  VW+K  NA  +++ ND+   +E+ +  L+ A P GMK+ + S+
Sbjct: 3   ITLARIDDRLIHGQVTTVWSKVANAQRIIICNDDVFNDEVRRTLLRQAAPPGMKVNVVSL 62

Query: 62  EESIALFKDPRATDKRIFVIVNSVKDACTIAKNITDLEAVNVANVGRFDKSDPATKVKLT 121
            E+++A++ +P+  D+ +F +  +  D  T+ +    +  +N+  +     +    K +LT
Sbjct: 63  EKAVAVYHNPQYQDETVFYLFTNPHDVLTMVRQGVQIATLNIGGM-----AWRPGKKQLT 117

Query: 122 SSLLLNTEELEAAKELASL-PDLDVFNQVLPSNTKVNL 158
             ++ L+ ++++A +EL  L   LD+  +V+ S+  VN+
Sbjct: 118 KAVSLDPQDIQAFRELDKLGVKLDL--RVVASDPSVNI 153
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1639> which encodes the amino acid sequence <SEQ ID 1640>. Analysis of this protein sequence reveals the following:

```
Possible site: 36
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1874(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 145/162 (89%), Positives = 152/162 (93%)

Query: 1   MITQIRVDDRLIHGQVAVVWTKELNAPLLVVANDEAAKNEITQMTLKMAVPNGMKLLIRS 60
            MITQIRVDDRLIHGQVAVVWTKELNAPLLVVANDEAAKNEITQMTLKMAVPNGMKLLIRS
Sbjct: 1   MITQIRVDDRLIHGQVAVVWTKELNAPLLVVANDEAAKNEITQMTLKMAVPNGMKLLIRS 60

Query: 61  VEESIALFKDPRATDKRIFVIVNSVKDACTIAKNITDLEAVNVANVGRFDKSDPATKVKL 120
            VE+SI LF DPRA DKRIFVIVNSVKDAC IAK + DLEAVNVANVGRFDKSDPA+KVK+
Sbjct: 61  VEDSIKLFNDPRAKDKRIFVIVNSVKDACAIAKEVPDLEAVNVANVGRFDKSDPASKVKV 120

Query: 121 TSSLLLNTEELEAAKELASLPDLDVFNQVLPSNTKVNLSQLV 162
            T SLLLN EE+ AAKEL SLP+LDVFNQVLPSNTKV+LSQLV
Sbjct: 121 TPSLLLNPEEMAAAKELVSLPELDVFNQVLPSNTKVHLSQLV 162
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 512

A DNA sequence (GBSx0550) was identified in *S.agalactiae* <SEQ ID 1641> which encodes the amino acid sequence <SEQ ID 1642>. Analysis of this protein sequence reveals the following:

```
Possible site: 46
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.22    Transmembrane    87 - 103 (   87 - 104)

----- Final Results -----
               bacterial membrane --- Certainty=0.1489(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1643> which encodes the amino acid sequence <SEQ ID 1644>. Analysis of this protein sequence reveals the following:

```
        Possible site: 33

   >>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -1.44   Transmembrane   87 - 103 (  87 - 104)

   ----- Final Results -----
                 bacterial membrane --- Certainty=0.1574(Affirmative) < succ>
                  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
                bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

An alignment of the GAS and GBS proteins is shown below:

```
    Identities = 115/141 (81%), Positives = 125/141 (88%)

   Query: 1    MKRKFLIGSHGKLASGLQSSIDILTGKGQEIQTIDAYIDDSDYTKSIVEFIDEIAPDEQG 60
               MKRKFLIGSHG+LASGLQSSIDIL G GQ ++TIDAY+DDSDYT  I +FI  +A DEQG
   Sbjct: 1    MKRKFLIGSHGRLASGLQSSIDILAGMGQALETIDAYVDDSDYTSQIDDFIAGVAADEQG 60

   Query: 61   LIFTDLLGGSVNQKMATAVMNSGKNNIFLITNSNLATLLSLLFLKPEEELTKEEIVTVIN 120
               LIFTDLLGGSVNQKM TAVMNSGK+NIFLITNSNLATLLSL+FLKP E LTK+EIVTVIN
   Sbjct: 61   LIFTDLLGGSVNQKMVTAVMNSGKDNIFLITNSNLATLLSLVFLKPGEALTKDEIVTVIN 120

   Query: 121  ESQVQLVDLSFKAGSEDDFFD 141
               ESQVQLVDL  +  SEDDFFD
   Sbjct: 121  ESQVQLVDLVPETNSEDDFFD 141
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 513

A DNA sequence (GBSx0551) was identified in *S.agalactiae* <SEQ ID 1645> which encodes the amino acid sequence <SEQ ID 1646>. Analysis of this protein sequence reveals the following:

```
   Possible site: 25
   >>> Seems to have no N-terminal signal sequence

   ----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2469(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 514

A DNA sequence (GBSx0552) was identified in *S.agalactiae* <SEQ ID 1647> which encodes the amino acid sequence <SEQ ID 1648>. This protein is predicted to be racemase. Analysis of this protein sequence reveals the following:

```
   Possible site: 41
   >>> Seems to have no N-terminal signal sequence
```

```
   INTEGRAL    Likelihood = -8.65   Transmembrane  319 - 335 ( 316 - 339)
   INTEGRAL    Likelihood = -6.10   Transmembrane   18 -  34 (  17 -  37)
   INTEGRAL    Likelihood = -5.68   Transmembrane  230 - 246 ( 227 - 248)
   INTEGRAL    Likelihood = -3.98   Transmembrane  254 - 270 ( 254 - 271)
   INTEGRAL    Likelihood = -3.56   Transmembrane  110 - 126 ( 110 - 129)
   INTEGRAL    Likelihood = -3.19   Transmembrane  161 - 177 ( 156 - 177)
   INTEGRAL    Likelihood = -1.97   Transmembrane  132 - 148 ( 132 - 153)
   INTEGRAL    Likelihood = -1.33   Transmembrane  286 - 302 ( 286 - 302)
   INTEGRAL    Likelihood = -0.59   Transmembrane   53 -  69 (  52 -  69)

----- Final Results -----
               bacterial membrane --- Certainty=0.4461(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAF71283 GB:AF253562 racemase [Enterococcus faecalis]
 Identities = 78/262 (29%), Positives = 129/262 (48%), Gaps = 29/262 (11%)

Query: 13  KQHNTSMISLLQYLFSILVILVHSGRLFS-QDVIHFTFKSFLGRMAVPYFLICTAFFLRG 71
            K + S I   +++ ++L++ +H+  LFS  +  +F F   +  +AVP+F + + FFL
Sbjct: 3   KNESYSGIDYFRFIAALLIVAIHTSPLFSFSETGNFIFTRIVAPVAVPFFFMTSGFFL-- 60

Query: 72  RIQQGLCNHSYFRKLIKK----YSMWTIIYLPY----GYFFFESLNIAKIYLLPGFIVAF 123
             I +  CN       IKK    Y +  ++Y+P     GYF  ++L       LP  I
Sbjct: 61  -ISRYTCNAEKLGAFIKKTTLIYGVAILLYIPINVYNGYFKMDNL-------LPNIIKDI 112

Query: 124 LYLGMSHTLWYIPAVILGWVIIQGLLKYVGTRGTFITVVVLYCIGAV-ETYSVFIQSTKF 182
            ++ G  + LWY+PA I+G  I   L+K V  R  F+   +LY IG   ++Y   ++S
Sbjct: 113 VFDGTLYHLWYLPASIIGAAIAWYLVKKVHYRKAFLIASILYIIGLFGDSYYGIVKSVSC 172

Query: 183 YPLMSTYMSIFQT---TRNGLFYTPVYLLAGYLLYDYFNTDLFTKSRGLK-YILFLLLLA 238
              L   Y  IFQ    TRNG+F+ P++ + G  + D  + + + K   ++ Y LF L+
Sbjct: 173 --LNVFYNLIFQLTDYTRNGIFFAPIFFVLGGYISD--SPNRYRKKNYIRIYSLFCLMFG 228

Query: 239 LENVLIYFN-QGLDKNFFLLAP 259
                L +F+ Q  D  + LL P
Sbjct: 229 KTLTLQHFDIQKHDSMYVLLLP 250
```

No corresponding DNA sequence was identified in *S.pyogenes*.

A related GBS gene <SEQ ID 8589> and protein <SEQ ID 8590> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 7
McG: Discrim Score:      0.23
GvH: Signal Score (-7.5): -5.77
     Possible site: 34
>>> Seems to have an uncleavable N-term signal seq
ALOM program   count: 3 value:  -5.68 threshold:  0.0
   INTEGRAL    Likelihood = -5.68   Transmembrane   41 -  57 (  38 -  59)
   INTEGRAL    Likelihood = -3.98   Transmembrane   65 -  81 (  65 -  82)
   INTEGRAL    Likelihood = -1.33   Transmembrane   97 - 113 (  97 - 113)
   PERIPHERAL  Likelihood =  5.78      10
 modified ALOM score:   1.64

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.3272(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS gene <SEQ ID 8591> and protein <SEQ ID 8592> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 5
McG: Discrim Score:     11.50
GvH: Signal Score (-7.5): -2.69
     Possible site: 32
>>> Seems to have an uncleavable N-term signal seq
ALOM program   count: 9 value:  -8.65 threshold:  0.0
   INTEGRAL    Likelihood = -8.65   Transmembrane  310 - 326 ( 307 - 330)
   INTEGRAL    Likelihood = -6.10   Transmembrane    9 -  25 (   8 -  28)
   INTEGRAL    Likelihood = -5.68   Transmembrane  221 - 237 ( 218 - 239)
   INTEGRAL    Likelihood = -3.98   Transmembrane  245 - 261 ( 245 - 262)
   INTEGRAL    Likelihood = -3.56   Transmembrane  101 - 117 ( 101 - 120)
   INTEGRAL    Likelihood = -3.19   Transmembrane  152 - 168 ( 147 - 168)
   INTEGRAL    Likelihood = -1.97   Transmembrane  123 - 139 ( 123 - 144)
   INTEGRAL    Likelihood = -1.33   Transmembrane  277 - 293 ( 277 - 293)
   INTEGRAL    Likelihood = -0.59   Transmembrane   44 -  60 (  43 -  60)
   PERIPHERAL  Likelihood =  5.78     190
 modified ALOM score:   2.23

*** Reasoning Step: 3

----- Final Results -----
              bacterial membrane --- Certainty=0.4461(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
ORF00153(307 - 1140 of 1632)
GP|7960293|gb|AAF71283.1|AF253562_7|AF253562(2  -  284  of  711)  racemase  {Enterococcus
faecalis}
%Match = 8.5
%Identity = 32.7  %Similarity = 54.0
Matches = 91  Mismatches = 113  Conservative Sub.s = 59

150       180       210       240       270       300       330       360
CEISFFIS*YG**GINNNYQIPFKAFQ*LFGIIEIFF*RDWYHSNDNL*KVMLRMKRSQCVDNKQHNTSMISLLQYLFSI
                                                               |  : | |  :::: ::
                                                             MTKNESYSGIDYFRFIAAL
                                                                     10

390       417       447       477       507       537       555
LVILVHSGRLFS-QDVIHFTFKSFLGRMAVPYFLICTAFFLRGRIQQGLCNHSYFRKLIKK----YSMWTIIYLP---Y-
|:: :|:  |||  :  :| |   :  :|||:|:: : |||   | :  ||   :  :|||    | :  ::|:|   |
LIVAIHTSPLFSFSETGNFIFTRIVAPVAVPFFFMTSGFFL---ISRYTCNAEKLGAFIKKTTLIYGVAILLYIPINVYN
         30        40        50        60           70        80        90

603       633       663       693       723       753       783       810
GYFFFESLNIAKIYLLPGFIVAFLYLGMSHTLWYIPAVILGWVIIQGLLKYVGTRGTFITVVVLYCIGAV-ETYSVFIQS
|||  ::|       ||  |   :: |  : |||:|| |:|  |   |:| |  |  |:   :|| ||   ::|   ::|
GYFKMDNL-------LPNIIKDIVFDGTLYHLWYLPASIIGAAIAWYLVKKVHYRKAFLIASILYIIGLFGDSYYGIVKS
                  110       120       130       140       150       160

840                 891       921       951       978      1008      1035
TKFYPLMSTYMSIFQTT---RNGLFYTPVYLLAGYLLYDYFNTDLFTKSRGLK-YILFLLLLALENVLIYFN-QGLDKNF
     |   |  ||| |   |||:|: |:::: |  : |  |   : |   :: | || |::     | :|: |  |  :
--VSCLNVFYNLIFQLTDYTRNGIFFAPIFFVLGGYISDSPNR--YRKKNYIRIYSLFCLMFGKTLTLQHFDIQKHDSMY
          180       190       200       210         220       230       240

1053      1080      1110      1140      1170      1200      1230      1260
FLLAP----LCAVFL-FNWSIRTSLFKEYRLSPLKQLSVYYFFLPPLFIGIVSYCLKSTSLVAHHQGKVIFVVTLALTHA
 || |    |  ::| |    || |  :   |  | : |||
VLLLPSVWCLFNLLLHFRGKRRTGL-RTISLDQLYHSSVYDCCNTIVCAELLHLQSLLVENSLVHYIAVCFASVVLAVVI
            260       270        280       290       300       310       320
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 515

A DNA sequence (GBSx0553) was identified in *S.agalactiae* <SEQ ID 1649> which encodes the amino acid sequence <SEQ ID 1650>. Analysis of this protein sequence reveals the following:

```
Possible site: 43
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3088(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 516

A DNA sequence (GBSx0554) was identified in *S.agalactiae* <SEQ ID 1651> which encodes the amino acid sequence <SEQ ID 1652>. Analysis of this protein sequence reveals the following:

```
Possible site: 35
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1446(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 517

A DNA sequence (GBSx0555) was identified in *S.agalactiae* <SEQ ID 1653> which encodes the amino acid sequence <SEQ ID 1654>. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 10
McG: Discrim Score:      8.28
GvH: Signal Score (-7.5): -2.11
     Possible site: 20
>>> Seems to have a cleavable N-term signal seq.
ALOM program   count: 6 value:  -8.33 threshold:  0.0
   INTEGRAL    Likelihood = -8.33   Transmembrane  358 - 374 ( 354 - 376)
   INTEGRAL    Likelihood = -8.23   Transmembrane  264 - 280 ( 257 - 290)
   INTEGRAL    Likelihood = -6.37   Transmembrane  210 - 226 ( 206 - 232)
   INTEGRAL    Likelihood = -5.95   Transmembrane  163 - 179 ( 160 - 180)
   INTEGRAL    Likelihood = -5.10   Transmembrane   23 -  39 (  21 -  40)
   INTEGRAL    Likelihood = -1.70   Transmembrane  297 - 313 ( 296 - 314)
   PERIPHERAL  Likelihood =  1.75     322
 modified ALOM score:   2.17

*** Reasoning Step: 3

----- Final Results -----
```

```
     bacterial membrane --- Certainty=0.4333(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
    bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 518

A DNA sequence (GBSx0556) was identified in *S.agalactiae* <SEQ ID 1655> which encodes the amino acid sequence <SEQ ID 1656>. This protein is predicted to be ABC transporter (ATP-bindingprot). Analysis of this protein sequence reveals the following:

```
Possible site: 59
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1510(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10199> which encodes amino acid sequence <SEQ ID 10200> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB88481 GB:AL353816 putative ABC transport system ATP-binding
           protein [Streptomyces coelicolor A3(2)]
 Identities = 104/284 (36%), Positives = 159/284 (55%), Gaps = 18/284 (6%)

Query: 6   TMLLQLDNITKSYGKKIVLNQISYQFTPGLYGLLGANGTGKTTLLNLMSHFTLADSGNIY 65
            T  +    ++  YG+   L+ +S + TPG+ GLLG NG GKTTLL +++    AD G
Sbjct: 2   TPTVSASGLSLHYGRTRALDDVSLRLTPGVTGLLGPNGAGKTTLLRVLATAVPADRGAFT 61

Query: 66  WNGQEQS-----EEFYRHIGFLPQHFRYYDQFTGIAFLNYIATLKGV-DKKKAKQEIPRL 119
              G +       +E  R +G+LPQ   ++  FT   F++Y+A LK + D+++  +E+ R+
Sbjct: 62  VLGHDPGSSRGRQEVRRRLGYLPQTPGFHPDFTAFEFVDYVAILKELADRRERHREVRRV 121

Query: 120 LELVGLGDVGKKKISSYSGGMKQRLGIAQALINDPEILILDEPTVGLDPKERVKFRHILS 179
            LE V LG+V  ++I   SGGM+QR+ +A AL+ DP  L+LDEPTVGLDP++R++FR +++
Sbjct: 122 LEEVDLGEVRGRRIKKLSGGMRQRVALAAALVGDPGFLVLDEPTVGLDPEQRMRFRELIA 181

Query: 180 QLSTNKIIILSTHIVSDVEAVAKEIIVLKNGKFIEHGNTAQLLKTIEGKVWEIT-TEPGL 238
                 + ++LSTH   DV  +   +IV+  G     G  A+L     G+VW  T  +PG
Sbjct: 182 GAGEGRTVLLSTHQTEDVAMLCHRVIVMAAGAVRFDGTPAELTARAAGRVWSSTEKDPG- 240

Query: 239 SQIPNIAIVNEKVFSDSRVFRVVSDICPSDSAQLVVPTLEDFYI 282
                  A    +  + S  FR V D  P   A+   PTLED Y+
Sbjct: 241 ------AKAGWRTGTGS--FRNVGD--PPPGAEPAEPTLEDGYL 274
```

There is also homology to SEQ ID 686.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 519

A DNA sequence (GBSx0557) was identified in *S.agalactiae* <SEQ ID 1657> which encodes the amino acid sequence <SEQ ID 1658>. This protein is predicted to be response regulator. Analysis of this protein sequence reveals the following:

```
Possible site: 38
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3781(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAC10170 GB:AJ278301 response regulator [Streptococcus pneumoniae]
 Identities = 136/242 (56%), Positives = 183/242 (75%)

Query: 1   MNIFILEDDFVQQAHFEKIIKEIRVQYNLHFKTVETFAKPVQLLESIYEIGLHNLFFLDI 60
           M IF+LEDDF QQ   E  I+++  ++++   + E F KP QLL  ++E G H LFFLDI
Sbjct: 1   MRIFVLEDDFSQQTRIETTIEKLLKEHHITLSSFEVFGKPDQLLAEVHEKGAHQLFFLDI 60

Query: 61  EIKNDEQMGLEVAKQIRQVDPYAQIVFVTTHSELMPLTFRYQVSALDYIDKGLSQEEFSQ 120
           EI+N+E  GLEVA++IR+ DPYA IVFVTTHSE MPL+FRYQVSALDYIDK LS EEF
Sbjct: 61  EIRNEEMKGLEVARKIREQDPYALIVFVTTHSEFMPLSFRYQVSALDYIDKALSAEEFES 120

Query: 121 RIEEVLLYVDGICNKPLVENSFYFKSRYSQVQLPFNDLLYIETSSRSHRVVLYTEKDRME 180
           RIE  LLY +   +K L E+ FYFKS+++Q Q PF ++  Y+ETS R HRV+LYT+ DR+E
Sbjct: 121 RIETALLYANSQDSKSLAEDCFYFKSKFAQFQYPFKEVYYLETSPRPHRVILYTKTDRLE 180

Query: 181 FTATLGDILKQEPRLFQCHRSFLVNPLNIFKVDRIDRLVYFQNGTTCLVSRNKVRDIVSI 240
           FTA+L ++ KQEPRL QCHRSFL+NP N+  +D+ ++L++F NG +CL++R KVR++
Sbjct: 181 FTASLEEVFKQEPRLLQCHRSFLINPANVVHLDKKEKLLFFPNGGSCLIARYKVREVSEA 240

Query: 241 VD 242
           ++
Sbjct: 241 IN 242
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1659> which encodes the amino acid sequence <SEQ ID 1660>. Analysis of this protein sequence reveals the following:

```
Possible site: 44
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2098(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 106/235 (45%), Positives = 159/235 (67%)

Query: 1   MNIFILEDDFVQQAHFEKIIKEIRVQYNLHFKTVETFAKPVQLLESIYEIGLHNLFFLDI 60
           MNIFILEDDF+QQ   E I+  I  +  +    +E F+ P +L ESI E G H L+FLDI
Sbjct: 2   MNIFILEDDFIQQTRIESIVVGILKETRIPCNQLEVFSTPQKLFESIQERGDHQLYFLDI 61

Query: 61  EIKNDEQMGLEVAKQIRQVDPYAQIVFVTTHSELMPLTFRYQVSALDYIDKGLSQEEFSQ 120
           EI    + GLE+A  IRQ DP A IVFVTTHSE  P++F+Y+VSALD+IDK   Q++F +
Sbjct: 62  EIGEYTRCGLELAAAIRQKDPNAVIVFVTTHSEFAPISFKYKVSALDFIDKAGGQKQFKE 121

Query: 121 RIEEVLLYVDGICNKPLVENSFYFKSRYSQVQLPFNDLLYIETSSRSHRVVLYTEKDRME 180
           +IEE + Y   + +    ++ F F++  ++++LP+ D+LY  T++  H+V L+T+ +R+E
Sbjct: 122 QIEECIRYTYDMMSSRESKDMFLFETPQTRLKLPYKDILYFATATTPHKVCLWTQTERLE 181
```

```
Query: 181 FTATLGDILKQEPRLFQCHRSFLVNPLNIFKVDRIDRLVYFQNGTTCLVSRNKVR 235
           F   L +I    P+LF CHRS+LVN   + ++D+  +L+YF+NG +C+VSR K++
Sbjct: 182 FYGNLSEIQAVAPKLFLCHRSYLVNLDKVVRIDKSKQLLYFENGDSCMVSRLKMK 236
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 520

A DNA sequence (GBSx0558) was identified in *S.agalactiae* <SEQ ID 1661> which encodes the amino acid sequence <SEQ ID 1662>. Analysis of this protein sequence reveals the following:

```
Possible site: 57
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2651(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1663> which encodes the amino acid sequence <SEQ ID 1664>. Analysis of this protein sequence reveals the following:

```
Possible site: 25
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0535(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 177/269 (65%), Positives = 219/269 (80%)

Query: 6   MAKCLTLNTHSWMEVNALKKLFDLAEHIFREKYDIICLQEVNQSISSPLAKSSPNYHPIE 65
           M K LTLNTHSWM+ N LKKL  LAEHI  EKYDIICLQE+NQ I S LA   P Y  +
Sbjct: 1   MTKVLTLNTHSWMQANTLKKLVALAEHILAEKYDIICLQEINQLIESELATDLPRYQALS 60

Query: 66  GTPALHQDNFALQLVHYLNLQGLHYHWTWAYNHIGYSKYHEGVAILSLKPLKPEDILVSA 125
           GTP++H+D+FAL L+HYL  +G HY+W+WAYNHIGY  Y EGVAILS +P+   DILVSA
Sbjct: 61  GTPSIHKDHFALLLIHYLQKRGQHYYWSWAYNHIGYDIYQEGVAILSKQPIHVSDILVSA 120

Query: 126 VDDETDYHTRRALVAETTLNDKVVTVVSLHFSWFEKGFAEEWKRLETTLLEVETPLLLMG 185
           +DDETDYHTRR+L+A+TTL+ K V VV++H SWF+KGF  EW++LE  LL +  PLLLMG
Sbjct: 121 MDDETDYHTRRSLIAKTTLDGKEVAVVNVHLSWFDKGFLGEWEKLEKELLTLNCPLLLMG 180

Query: 186 DFNNPTGNQGYELVLNSPLALKDSHQIANHVFGDHTIMADIDGWEGNKKALKVDHIFTSE 245
           DFNNPT   GY++++ SPL L+DSH+ A+HVFGDH+I+ADIDGW+GNK+ALKVDH+FTS+
Sbjct: 181 DFNNPTDQDGYQVMMGSPLDLQDSHKGADHVFGDHSIVADIDGWQGNKEALKVDHVFTSK 240

Query: 246 DLSISSSQVVFEGGEAPVVSDHYGLEITM 274
           D  I SS++ FEGG+APVVSDHYGLE+T+
Sbjct: 241 DFIIRSSKITFEGGDAPVVSDHYGLEVTL 269
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 521

A DNA sequence (GBSx0559) was identified in *S.agalactiae* <SEQ ID 1665> which encodes the amino acid sequence <SEQ ID 1666>. This protein is predicted to be PTS system, glucose-specific enzyme II, A component (ptsG). Analysis of this protein sequence reveals the following:

```
Possible site: 37
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -8.07   Transmembrane  193 - 209 ( 189 - 217)
   INTEGRAL    Likelihood = -7.86   Transmembrane   28 -  44 (  24 -  48)
   INTEGRAL    Likelihood = -6.48   Transmembrane  431 - 447 ( 421 - 449)
   INTEGRAL    Likelihood = -2.92   Transmembrane  153 - 169 ( 153 - 170)
   INTEGRAL    Likelihood = -2.81   Transmembrane   93 - 109 (  93 - 111)
   INTEGRAL    Likelihood = -2.39   Transmembrane  370 - 386 ( 370 - 388)
   INTEGRAL    Likelihood = -2.28   Transmembrane   68 -  84 (  68 -  84)

----- Final Results -----
               bacterial membrane --- Certainty=0.4227(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10201> which encodes amino acid sequence <SEQ ID 10202> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAD00281 GB:U78600 putative ptsG protein [Streptococcus mutans]
 Identities = 294/409 (71%), Positives = 342/409 (82%), Gaps = 7/409 (1%)

Query: 293 DLINLKGS-NSSQYHHLLTSVTPARFKVGQMIGASGILMGLSYAMYRNVDKDKKLKYKSM 351
           DLI+LKG+ + SQYHHLLTSVTPARFKVGQMIG+SGILMGL+ AMYRNVD DKK KYK M
Sbjct: 3   DLIHLKGAGHMSQYHHLLTSVTPARFKVGQMIGSSGILMGLTLAMYRNVDPDKKEKYKGM 62

Query: 352 FISAAAATFLTGVTEPIEYMFMFAAMPLYLVYAVVQGCAFAMADIVNLRVHSFGNIEFLT 411
           F+SAA A FLTGVTEP+EYMFMFAA+PLYLVYAVVQG AFA AD+++LRVHSFGNIEFLT
Sbjct: 63  FLSAAVAVFLTGVTEPLEYMFMFAALPLYLVYAVVQGLAFASADLIHLRVHSFGNIEFLT 122

Query: 412 RVPMGIKAGLGGDIFNFVWVTLLFAVLMYFIANFMIKKFNLATAGRNGNYDNEEVDNAPS 471
           + PM IKAGL  DI NF+ V+++F V MYFI NFMIKKFNLAT+GRNGNYD  + D +
Sbjct: 123 KTPMAIKAGLAMDIVNFIVVSVVFGVAMYFITNFMIKKFNLATSGRNGNYDTGD-DASDE 181

Query: 472 TAS----GSADANSQVVQVINLLGGRDNIEDVDACMTRLRVTVKDGNSVGSEAAWKKAGA 527
           TAS    G+A+ANSQ+V++INLLGG++NI DVDACMTRLR+TV D    VG EAAWKKAGA
Sbjct: 182 TASNSNAGTANANSQIVKIINLLGGKENISDVDACMTRLRITVTDVAKVGDEAAWKKAGA 241

Query: 528 MGLVLKGNGVQAIYGPKADVLKSDIQDLLDSGTVIPIVDLETGQPVAAAPVTTYKGITEE 587
           MGL++KGNGVQA+YGPKADVLKSDIQDLLDSG  IP  D+   +    A V ++KG+TEE
Sbjct: 242 MGLIVKGNGVQAVYGPKADVLKSDIQDLLDSGVDIPKTDVTAPEEDKTADV-SFKGVTEE 300

Query: 588 IVSVANGQVEALDVVKDPVFSQKMMGDGFAVEPTDGNIYVPVSGTVTSVFPTKHAFGLLT 647
           + +VA+GQV  +  V DPVFSQKMMGDGFAVEP +GNIY PV+G VTSVFPTKHA GLLT
Sbjct: 301 VATVADGQVLPITQVHDPVFSQKMMGDGFAVEPENGNIYSPVAGLVTSVFPTKHALGLLT 360

Query: 648 ESGLEVLVHIGLDTVALDGQPFEVKISSGQKVVAGDLAVVADLEAIKAA 696
           + GLEVLVH+GLDTVAL+G PF  K+  GQ+V  GDL +VADLEAIK+A
Sbjct: 361 DDGLEVLVHVGLDTVALNGAPFSAKVKDGQRVALGDLLLVADLEAIKSA 409
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1667> which encodes the amino acid sequence <SEQ ID 1668>. Analysis of this protein sequence reveals the following:

```
     Possible site: 33
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood =-13.43   Transmembrane  186 - 202 ( 181 - 213)
   INTEGRAL    Likelihood = -6.79   Transmembrane  419 - 435 ( 412 - 442)
   INTEGRAL    Likelihood = -5.52   Transmembrane   61 -  77 (  57 -  82)
```

```
    INTEGRAL    Likelihood = -3.56   Transmembrane  363 - 379 ( 363 - 381)
    INTEGRAL    Likelihood = -1.97   Transmembrane  143 - 159 ( 142 - 160)
    INTEGRAL    Likelihood = -0.16   Transmembrane  343 - 359 ( 343 - 359)

 ----- Final Results -----
              bacterial membrane --- Certainty=0.6371(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAD00281 GB:U78600 putative ptsG protein [Streptococcus mutans]
 Identities = 288/407 (70%), Positives = 331/407 (80%), Gaps = 2/407 (0%)

Query: 286 DLVHLKGSD-ASAYSHLMDSVTPARFKVGQMIGATGTLMGVALAMYRNVDADKKHTYKMM 344
            DL+HLKG+   S Y HL+ SVTPARFKVGQMIG++G LMG+ LAMYRNVD DKK  YK M
Sbjct: 3   DLIHLKGAGHMSQYHHLLTSVTPARFKVGQMIGSSGILMGLTLAMYRNVDPDKKEKYKGM 62

Query: 345 FISAAAAVFLTGVTEPLEYLFMFAAMPLYIVYALVQGASFAMADLVNLRVHSFGNIELLT 404
            F+SAA AVFLTGVTEPLEY+FMFAA+PLY+VYA+VQG +FA ADL++LRVHSFGNIE LT
Sbjct: 63  FLSAAVAVFLTGVTEPLEYMFMFAALPLYLVYAVVQGLAFASADLIHLRVHSFGNIEFLT 122

Query: 405 RTPMALKAGLGMDVINFVWVSVLFAVIMYFIADMMIKKMHLATAGRLGNYDA-DILGDRN 463
            +TPMA+KAGL MD++NF+ VSV+F V MYFI + MIKK +LAT+GR GNYD  D   D
Sbjct: 123 KTPMAIKAGLAMDIVNFIVVSVVFGVAMYFITNFMIKKFNLATSGRNGNYDTGDDASDET 182

Query: 464 TQTRPTQVADSNSQVVQIVNLLGGAGNIDDVDACMTRLRVTVKDPAKVGAEDDWKKAGAI 523
                    A++NSQ+V+I+NLLGG  NI DVDACMTRLR+TV D AKVG E  WKKAGA+
Sbjct: 183 ASNSNAGTANANSQIVKIINLLGGKENISDVDACMTRLRITVTDVAKVGDEAAWKKAGAM 242

Query: 524 GLIQKGNGVQAVYGPKADILKSDIQDLLDSGALIPEVNMSQLTSKPTPAKDFKHVTEDVL 583
            GLI KGNGVQAVYGPKAD+LKSDIQDLLDSG  IP+ +++      T    FK VTE+V
Sbjct: 243 GLIVKGNGVQAVYGPKADVLKSDIQDLLDSGVDIPKTDVTAPEEDKTADVSFKGVTEEVA 302

Query: 584 SVADGMVLPITGVKDQVFAAKMMGDGFAVEPTHGNIYAPVAGLVTSVFPTKHAFGLLTDN 643
            +VADG VLPIT V D VF+ KMMGDGFAVEP +GNIY+PVAGLVTSVFPTKHA GLLTD+
Sbjct: 303 TVADGQVLPITQVHDPVFSQKMMGDGFAVEPENGNIYSPVAGLVTSVFPTKHALGLLTDD 362

Query: 644 GLEVLVHVGLDTVALNGVPFSVKVSEGQRVHAGDLLVVADLAAIKSA 690
            GLEVLVHVGLDTVALNG PFS KV +GQRV  GDLL+VADL AIKSA
Sbjct: 363 GLEVLVHVGLDTVALNGAPFSAKVKDGQRVALGDLLLVADLEAIKSA 409
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 517/731 (70%), Positives = 606/731 (82%), Gaps = 7/731 (0%)

Query: 8   MKNNVKQLFSFEFWQKFGKALMVVIAVMPAAGLMVSIGNSISLLDPSNVLLGRIANVIAQ 67
            MK + KQLF FEFWQKFGK LMVVIAVMPAAGLM+SIGNSI +++  +  L  + N+IAQ
Sbjct: 1   MKTSFKQLFRFEFWQKFGKCLMVVIAVMPAAGLMISIGNSIPMINHDSAFLASLGNIIAQ 60

Query: 68  IGWGVIGNLHILFALAIGGSWAKERAGGAFAAGLSFILINLITGNFFGVKTDMLADSKAT 127
            IGW VI NLH+LFALAIGGSWAKERAGGAFA+GL+F+LIN ITG F+GV + MLAD +A
Sbjct: 61  IGWAVIVNLHLLFALAIGGSWAKERAGGAFASGLAFVLINRITGAFYGVSSTMLADPEAK 120

Query: 128 VQTVFGATIRVSDYFVNVLGQPALNMGVFVGIISGFVGATAFNKYYNYRKLPDALTFFNG 187
            + ++ G  + V DYF +VL  PALN GVFVGII+GFVGATA+NKYYNYRKLP+ LTFFNG
Sbjct: 121 ITSLLGTQMIVKDYFTSVLESPALNTGVFVGIIAGFVGATAYNKYYNYRKLPEVLTFFNG 180

Query: 188 KRFVPFVVIYRSVIVALILSVFWPVVQSGINGFGKWIASSQDSAPILAPFVYGTLERLLL 247
            KRFVPFVVI RS+ VALIL V WPV+QSGIN FG WIASSQDSAPILAPF+YGTLERLLL
Sbjct: 181 KRFVPFVVILRSIFVALILVVVWPVIQSGINSFGMWIASSQDSAPILAPFLYGTLERLLL 240

Query: 248 PFGLHHMLTIPMNYTQLGGTYTVLTGATKGAQVLGQDPLWLAWVGDLINLKGSNSSQYHH 307
            PFGLHHMLTIPMNYT LGGTY V+TGA  G +V GQDPLWLAWV DL++LKGS++S Y H
Sbjct: 241 PFGLHHMLTIPMNYTALGGTYEVMTGAAAGTKVFGQDPLWLAWVTDLVHLKGSDASAYSH 300

Query: 308 LLTSVTPARFKVGQMIGASGILMGLSYAMYRNVDKDKKLKYKSMFISAAAATFLTGVTEP 367
            L+ SVTPARFKVGQMIGA+G LMG++ AMYRNVD DKK  YK MFISAAAA FLTGVTEP
```

```
Sbjct: 301 LMDSVTPARFKVGQMIGATGTLMGVALAMYRNVDADKKHTYKMMFISAAAAVFLTGVTEP 360

Query: 368 IEYMFMFAAMPLYLVYAVVQGCAFAMADIVNLRVHSFGNIEFLTRVPMGIKAGLGGDIFN 427
            +EY+FMFAAMPLY+VYA+VQG +FAMAD+VNLRVHSFGNIE LTR PM +KAGLG D+ N
Sbjct: 361 LEYLFMFAAMPLYIVYALVQGASFAMADLVNLRVHSFGNIELLTRTPMALKAGLGMDVIN 420

Query: 428 FVWVTLLFAVLMYFIANFMIKKFNLATAGRNGNYDNEEVD--NAPSTASGSADANSQVVQ 485
            FVWV++LFAV+MYFIA+ MIKK +LATAGR GNYD + +   N  +  +  AD+NSQVVQ
Sbjct: 421 FVWVSVLFAVIMYFIADMMIKKMHLATAGRLGNYDADILGDRNTQTRPTQVADSNSQVVQ 480

Query: 486 VINLLGGRDNIEDVDACMTRLRVTVKDGNSVGSEAAWKKAGAMGLVLKGNGVQAIYGPKA 545
            ++NLLGG  NI+DVDACMTRLRVTVKD   VG+E  WKKAGA+GL+ KGNGVQA+YGPKA
Sbjct: 481 IVNLLGGAGNIDDVDACMTRLRVTVKDPAKVGAEDDWKKAGAIGLIQKGNGVQAVYGPKA 540

Query: 546 DVLKSDIQDLLDSGTVIPIVDLE--TGQPVAAAPVTTYKGITEEIVSVANGQVEALDVVK 603
            D+LKSDIQDLLDSG +IP V++   T +P    P   +K +TE+++SVA+G V  +  VK
Sbjct: 541 DILKSDIQDLLDSGALIPEVNMSQLTSKP---TPAKDFKHVTEDVLSVADGMVLPITGVK 597

Query: 604 DPVFSQKMMGDGFAVEPTDGNIYVPVSGTVTSVFPTKHAFGLLTESGLEVLVHIGLDTVA 663
            D VF+ KMMGDGFAVEPT GNIY PV+G VTSVFPTKHAFGLLT++GLEVLVH+GLDTVA
Sbjct: 598 DQVFAAKMMGDGFAVEPTHGNIYAPVAGLVTSVFPTKHAFGLLTDNGLEVLVHVGLDTVA 657

Query: 664 LDGQPFEVKISSGQKVVAGDLAVVADLEAIKAAGKETSVIIVFTNVSDIKTVKLEKSGPQ 723
            L+G PF VK+S GQ+V AGDL VVADL AIK+A +ET +++ FTN ++I+ V L   G Q
Sbjct: 658 LNGVPFSVKVSEGQRVHAGDLLVVADLAAIKSAERETIIVVAFTNTTEIQDVTLTSLGAQ 717

Query: 724 IAKTVVAKVEL 734
             AKT VA VEL
Sbjct: 718 PAKTKVATVEL 728
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 522

A DNA sequence (GBSx0560) was identified in *S.agalactiae* <SEQ ID 1669> which encodes the amino acid sequence <SEQ ID 1670>. Analysis of this protein sequence reveals the following:

```
Possible site: 14
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2266(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 523

A DNA sequence (GBSx0561) was identified in *S.agalactiae* <SEQ ID 1671> which encodes the amino acid sequence <SEQ ID 1672>. This protein is predicted to be alkaline phosphatase synthesis sensor protein phor (hpkA). Analysis of this protein sequence reveals the following:

```
Possible site: 34
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-13.96    Transmembrane  160 - 176 ( 148 - 183)
   INTEGRAL    Likelihood = -8.65    Transmembrane   20 -  36 (  13 -  41)
```

```
----- Final Results -----
               bacterial membrane --- Certainty=0.6583(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 8595> which encodes amino acid sequence <SEQ ID 8596> was also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 6
SRCFLG: 0
McG: Length  of  UR:   26
     Peak Value of UR:   3.27
     Net Charge of CR: 3
McG: Discrim Score:     14.63
GvH: Signal Score (-7.5): -5.64
     Possible site: 26
>>> Seems to have an uncleavable N-term signal seq
Amino Acid Composition: calculated from 1
ALOM program   count: 2 value: -13.96 threshold:  0.0
   INTEGRAL    Likelihood =-13.96   Transmembrane  152 - 168 ( 140 - 175)
   INTEGRAL    Likelihood = -8.65   Transmembrane   12 -  28 (   5 -  33)
   PERIPHERAL  Likelihood =  1.59     135
 modified ALOM score:   3.29
icml HYPID: 7  CFP: 0.658

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.6583(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS gene <SEQ ID 8593> and protein <SEQ ID 8594> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 6
McG: Discrim Score:     14.63
GvH: Signal Score (-7.5): -5.64
     Possible site: 26
>>> Seems to have an uncleavable N-term signal seq
ALOM program   count: 2 value: -13.96 threshold:  0.0
   INTEGRAL    Likelihood =-13.96   Transmembrane  152 - 168 ( 140 - 175)
   INTEGRAL    Likelihood = -8.65   Transmembrane   12 -  28 (   5 -  33)
   PERIPHERAL  Likelihood =  1.59     135
 modified ALOM score:   3.29

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.6583(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
34.9/61.1% over 363aa
                                             Thermotoga maritima
  EGAD|131465| sensor histidine kinase HpkA Insert characterized
  GP|1575578|gb|AAC44437.1||U67196 histidine protein kinase Insert characterized
  GP|4982228|gb|AAD36721.1|AE001807_12|AE001807   sensor   histidine   kinase   HpkA   Insert
characterized
  PIR|C72228|C72228 sensor histidine kinase HpkA - (strain MSB8) Insert characterized

ORF00680(919 - 1977 of 2277)
```

```
EGAD|131465|TM1654(48 - 411 of 412) sensor histidine kinase HpkA {Thermotoga maritima}
GP|1575578|gb|AAC44437.1||U67196 histidine protein kinase {Thermotoga maritima}
GP|4982228|gb|AAD36721.1|AE001807_12|AE001807 sensor histidine kinase HpkA {Thermotoga
maritima} PIR|C72228|C72228 sensor histidine kinase HpkA - Thermotoga maritima (strain
MSB8)
%Match = 13.6
%Identity = 34.8  %Similarity = 61.0
Matches = 125  Mismatches = 134  Conservative Sub.s = 94

720       750       780       810       840       870       900       930
AAQRLNNGTIVRLSVAQQTIFYLLLGMISPLAIIILLAIILSVLIARYIAKKVSEPLNNIDLDHPLSNDSYEEITPLLRR
                                 : ::  :| ||:     |::||    |:    :  :       |: |
                             MSVFLFVIVAVLFVLLFLVFKKRLSEYKILIEKLSDMLGEKGVPPLYLFER
                                     10        20        30        40        50

960       990       1020      1050      1080      1110      1140      1170
LDSHQAKIQHQKLLLQKRQKEFDTIISKIKEGMILLDDQARIVSINAEALKLFQINDDWHGRFMMEVSRDLTLKDLIDQG
|  :   ::     ::  :  | ||:: : | : :|| : :|  :|  | :| |   :  ||   |:  |  : :::::
LKKYVDNLKETISRVEVSRDNFLTILNSLSEPIFILDREGKITFLNEIARELVQGRINPEGRPYYEIFEDYYINEMVEET
                 70        80        90       100       110       120       130

1197      1215      1245      1275      1305      1335      1365      1395
LKGKK-KEAN----IGIENNHYRVLVRPTTDNNRVTGLVVLLFDVTDQLQMEQLQREFTANVSHELKTPLHVISGYSELL
:| :: :|      :| |  :: | | |    :    :|:|: ||| : ::::::||| | |||||:|||  | ||:| |
IKSEEPQEGTLVTYVGNEKKYFHVKVIPVELKSGDKIFVILFHDVTKERKLDEMRREFIATVSHELRTPLTSIHGYAETL
                150       160       170       180       190       200       210

1452      1482      1512      1539      1569      1599      1629
ANQMVPNEE-VPQFAAKIHKESERLVKLVEDIINLSHLDEQE-KLPQETVNLYDLTQKVLEGLQAKADKKHIQINFNGEE
    : |:| | :|   | :|| |: :|: |:::|  ::| |  :  : |:| :: : |   :|  |:: :  :    |:
LEDDLENKELVKRFLKIIEEESARMTRLINDLLDLEKIEESEANFEMKDVDLCEVIEYVYRIIQPIAEENEVDLIVECED
                230       240       250       260       270       280       290

1659      1689      1713                1767      1797      1827      1857
AILRGNPVLLNSLVYNLCDNAITYNH--EKGQVNVTLK--NSPDTITLEVSDTGLGIAEKDKKRIFERFYRVDKSRSKIV
 ::|||   |  :: || |||: |    |||:  | ::  ::|| : :|| ||| || :: : ||||:||||||:||: :
VVVRGNKERLIQMLLNLVDNAVKYTSLKEKGEKKVWVRAYDTPDWVVVEVEDTGPGIPKEAQSRIFEKFYRVDKARSRKM
                310       320       330       340       350       360       370

1887      1917      1947      1977      2007      2037      2067      2097
GGTGLGLSIVKSALDFHNGSIKVDSHLGQGTTMTVLLHKQ*KLTNKSLDDII*TFLVIQKKYISKGLQKTNKCYNKTXX*
|||||||:|||: :| | | |:|:| : ||| | ||| |:
GGTGLGLTIVKTIVDKHGGKIEVESEINQGTLMRVLLPKRR
                390       400       410
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAB06875 GB:AP001517 two-component sensor histidine kinase
           involved in phosphate regulation [Bacillus halodurans]
 Identities = 176/589 (29%), Positives = 315/589 (52%), Gaps = 47/589 (7%)

Query: 9   MTKKIFRTTLSASLGIVLVTILMIMG------------FLYNYFNHIQREQLRTQTALAS 56
           MTK  +R  L+    ++ VT+L++ G            +L N  + +++E     + + +
Sbjct: 1   MTKFRYRLVLA----VLTVTLLVMAGLGLVIGQIFKNVYLENLTDRLKKETYLAASMVEN 56

Query: 57  QGISF-EGKDYFENLKTS-NVRITWVDNKGQVLYDTQSDAKHMKNHANRQEIKEAIKSGY 114
           + + F E +   E +    + R+T +   G V+ ++ +D   M+NHA+R E  E ++ G
Sbjct: 57  EAVLFNEVQTLTEEISQKLDARVTIILADGTVVGESAADPAEMENHADRPEFTE-LEEGI 115

Query: 115 GESTRWSATL-TEKSIYAAQRLN--NGTI--VRLSVAQQTIFYLLLGMISPLAIIILLAI 169
               R+S T+ TE   YA    N  N TI  VRL +  + +  +   + + L +   +A
Sbjct: 116 ---VRYSTTVETELLFYAVPIQNEANETIGYVRLGLPIEAVNSVNRTLWAILIVSFTIAF 172

Query: 170 ILSVLIARYIAKKVSEPLNNI----------DLDHPLSNDSYEEITPLLRRLDSHQAKIQ 219
           ++ V +   IA ++  P+ +           D     S +S +E+  L R ++    ++
Sbjct: 173 LVIVSVTYRIANQMIRPIESATVVANKLAEGDYQARTSEESRDEVGQLNRSINVLAYNLE 232

Query: 220 HQKLLLQKRQKEFDTIISKIKEGMILLDDQARIVSINAEALKLFQINDD-WHGRFMMEVS 278
```

```
                 Q +++  +T+I  +  G+IL++ +  I  IN     +FQ + D W  +   +V
Sbjct: 233 QLTKRHQVQKERLETLIENMGSGLILINTRGDISLINKTCHDIFQEDTDLWLHQLYHDVI 292

Query: 279 RDLTLKDLIDQGLKGKKKEAN-----IGIENNHYRVLVRPTT-DNNRVTGLVVLLFDVTD 332
           +   +  ++      +K++       I +E  H+ V   P   +N ++ G+ ++  D+T+
Sbjct: 293 KHKEIIKIVQDIFLTEKRQRRQVKLPIHLEYRHFDVHGAPIVRENGKLKGIALVFHDITE 352

Query: 333 QLQMEQLQREFTANVSHELKTPLHVISGYSELLANQMVPNEEV-PQFAAKIHKESERLVK 391
             ++EQ++++F ANVSHELKTP+  I G++E L +  + +E++  QF   I KESERL
Sbjct: 353 LKKLEQVRKDFVANVSHELKTPVTSIKGFTETLLDGAMHDEQLRDQFLHIIWKESERLQS 412

Query: 392 LVEDIINLSHLDEQE-KLPQETVNLYDLTQKVLEGLQAKADKKHIQINFNGEEAI-LRGN 449
           L+ D++ LS +++   +L  +  NL+ +  +V+  L+ KA++K I I+ + E +  L G+
Sbjct: 413 LIHDLLELSKIEQNYFQLNWQQTNLFAVVSEVMTLLKGKAEEKGIDISLSAEGSFDLEGD 472

Query: 450 PVLLNSLVYNLCDNAITYNHEKGQVNVTLKNSPDTITLEVSDTGLGIAEKDKKRIFERFY 509
           P  L  +  NL +NAITY    G++++ LK+  D +  EV+DTG+GI E +  RIFERFY
Sbjct: 473 PERLKQIAINLVNNAITYTSNGGRIDLALKDHGDVVEFEVNDTGIGIRESEIPRIFERFY 532

Query: 510 RVDKSRSKIVGGTGLGLSIVKSALDFHNGSIKVDSHLGQGTTMTVLLHK 558
           RVD++RS+  GGTGLGL+IVK  ++ H G I V+S  G+GTT T+  H+
Sbjct: 533 RVDRARSRNSGGTGLGLAIVKHLVEAHQGKILVESEFGKGTTFTIQFHR 581
```

There is also homology to SEQ ID 1178.

SEQ ID 8594 (GBS340) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 173 (lane 10; MW 86kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 11 (lane 7; MW 61.5kDa) and in Figure 77 (lane 10; MW 62kDa).

Purified GBS340-GST is shown in Figure 223, lane 2; purified GBS340-His is shown in Fig. 191, lane 9.

The purified GBS340-GST fusion product was used to immunise mice. The resulting antiserum was used for Western blot (Figure 254A), FACS (Figure 254B), and in the *in vivo* passive protection assay (Table III). These tests confirm that the protein is immunoaccessible on GBS bacteria and that it is an effective protective immunogen.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 524

A DNA sequence (GBSx0562) was identified in *S.agalactiae* <SEQ ID 1673> which encodes the amino acid sequence <SEQ ID 1674>. This protein is predicted to be phosphate regulon transcriptional regulatory protein phob (phoB). Analysis of this protein sequence reveals the following:

```
Possible site: 28
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2617(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10203> which encodes amino acid sequence <SEQ ID 10204> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAC73502 GB:AE000146 positive response regulator for pho
```

```
          regulon, sensor is PhoR (or CreC) [Escherichia coli K12]
 Identities = 98/224 (43%), Positives = 138/224 (60%), Gaps = 2/224 (0%)

Query: 2   IYCVEDDADIREMMLYTLQMAGFKAQGFSSSELFWEAIQEKVPDLILLDIMLPGDDGLTI 61
           I  VED+A IREM+ + L+  GF+       +     + E  PDLILLD MLPG  G+
Sbjct: 5   ILVVEDEAPIREMVCFVLEQNGFQPVEAEDYDSAVNQLNEPWPDLILLDWMLPGGSGIQF 64

Query: 62  LERLRRKHQTEMIPVIMTTAKGSEYDKVKGLDLGADDYLVKPFGMMEMISRIKAVLRRSR 121
           ++ L+R+  T  IPV+M TA+G E D+V+GL+ GADDY+ KPF   E+++RIKAV+RR
Sbjct: 65  IKHLKRESMTRDIPVVMLTARGEEEDRVRGLETGADDYITKPFSPKELVARIKAVMRRIS 124

Query: 122 QVDSKAHIIIGNLEIDPTNYWVKRGTEKIHLTLKEFELLVLFFRNPNRVFTRQELLDKVW 181
            +  +  I +  L +DPT++ V  G E + +   EF+LL  F  +P RV++R++LL+ VW
Sbjct: 125 PMAVEEVIEMQGLSLDPTSHRVMAGEEPLEMGPTEFKLLHFFMTHPERVYSREQLLNHVW 184

Query: 182 GEQFLGETRTVDVHIGTLRTKLGEDGY--LIATVRGVGYRLEER 223
           G     E RTVDVHI  LR  L   G+  ++ TVRG GYR   R
Sbjct: 185 GTNVYVEDRTVDVHIRRLRKALEPGGHDRMVQTVRGTGYRFSTR 228
```

There is also homology to SEQ ID 1182.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 525

A DNA sequence (GBSx0563) was identified in *S.agalactiae* <SEQ ID 1675> which encodes the amino acid sequence <SEQ ID 1676>. This protein is predicted to be phosphate transport system regulatory protein (phoU). Analysis of this protein sequence reveals the following:

```
Possible site: 33
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1188(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAG08750 GB:AE004948 phosphate uptake regulatory protein PhoU
           [Pseudomonas aeruginosa]
 Identities = 66/213 (30%), Positives = 119/213 (54%), Gaps = 4/213 (1%)

Query: 2   IRSRFASQLNDLNKEIIFMGALCEDIIGKSLGALTNSNDVYLDDISETYHKIEQMERDIE 61
           I  +F ++L D+   ++ MG L E  +  ++ AL +++      + E   +I QMER+I+
Sbjct: 11  ISQQFNAELEDVRSHLLAMGGLVEKQVNDAVNALIDADSGLAQQVREIDDQINQMERNID 70

Query: 62  ERCLKLLLRQQPVAKDLRRISSALKMVYDMKRIGAQAYEIAEIVSLGHIIQGSGSERD-- 119
           E C+++L R+QP A DLR I S  K V D++RIG +A ++A       + +   S R
Sbjct: 71  EECVRILARRQPAASDLRLIISISKSVIDLERIGDEASKVARRAI--QLCEEGESPRGYV 128

Query: 120 QLNSMSNNVISMLTKSIDAFIYDNEEQAHQVIEQDRTVNQEFDTIKKQLVLYFSVQDVDG 179
           ++  + + V  M+ +++DAF   + + A  V + D+TV++E+ T  ++LV Y
Sbjct: 129 EVRHIGSQVQKMVQEALDAFARFDADLALSVAQYDKTVDREYKTALRELVTYMMEDPRAI 188

Query: 180 EYPIDVLMIAKYLERIGDHTVNIAKWVLFSITG 212
              ++++   + LERIGDH  NIA+ V++ + G
Sbjct: 189 SRVLNIIWALRSLERIGDHARNIAELVIYLVRG 221
```

There is also homology to SEQ ID 1678.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 526

A DNA sequence (GBSx0564) was identified in *S.agalactiae* <SEQ ID 1679> which encodes the amino acid sequence <SEQ ID 1680>. This protein is predicted to be ATP-binding cassette protein PstB (pstB-2). Analysis of this protein sequence reveals the following:

```
Possible site: 52
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2432(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10205> which encodes amino acid sequence <SEQ ID 10206> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAD22041 GB:AF118229 ATP-binding cassette protein PstB
           [Streptococcus pneumoniae]
 Identities = 166/245 (67%), Positives = 211/245 (85%), Gaps = 1/245 (0%)

Query: 10  INNLDLYYGEFHALKDVNLDIEEKEITAFIGPSGCGKSTLLKSINRMNDLVKNCKITGDI 69
            + +LDL+YG+F ALK++++ + E++ITA IGPSGCGKST LK++NRMNDLV +C I G +
Sbjct: 6   VRHLDLFYGDFQALKNISIQLPERQITALIGPSGCGKSTFLKTLNRMNDLVPSCHIEGQV 65

Query: 70  TLEGEDVYR-QLDINQLRKKVGMVFQKPNPFPMSIYDNVAFGPRTHGIHSKAELDDIVER 128
             L+ +D+Y  + ++NQLRK+VGMVFQ+PNPF MSIYDNVA+GPRTHGI  K +LD +VE+
Sbjct: 66  LLDEQDIYSSKFNLNQLRKRVGMVFQQPNPFAMSIYDNVAYGPRTHGIRDKKQLDALVEK 125

Query: 129 SLKQAALWDEVKDRLHKSALGMSGGQQQRLCIARALAIEPDVLLMDEPTSALDPISTAKI 188
            SLK AA+W+EVKD L KSA+ +SGGQQQRLCIARALA+EPD+LLMDEPTSALDPIST KI
Sbjct: 126 SLKGAAIWEEVKDDLKKSAMSLSGGQQQRLCIARALAVEPDILLMDEPTSALDPISTLKI 185

Query: 189 EELVIQLKKNYTIVIVTHNMQQAVRISDKTAFFLMGEVVEYNKTSQLFSLPQDERTENYI 248
            E+L+ QLKK+YTI+IVTHNMQQA RISDKTAFFL GE+ E+  T  +F+ P+D+RTE+YI
Sbjct: 186 EDLIQQLKKDYTIIIVTHNMQQASRISDKTAFFLTGEICEFGDTVDVFTNPKDQRTEDYI 245

Query: 249 TGRFG 253
            +GRFG
Sbjct: 246 SGRFG 250
```

There is also homology to SEQ ID 1682.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 527

A DNA sequence (GBSx0565) was identified in *S.agalactiae* <SEQ ID 1683> which encodes the amino acid sequence <SEQ ID 1684>. This protein is predicted to be transmembrane protein PstA (pstA-2). Analysis of this protein sequence reveals the following:

```
Possible site: 38
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood =-13.11    Transmembrane  265 - 281 ( 255 - 286)
   INTEGRAL    Likelihood = -8.81    Transmembrane   79 -  95 (  68 - 100)
   INTEGRAL    Likelihood = -4.78    Transmembrane  195 - 211 ( 192 - 213)
   INTEGRAL    Likelihood = -4.67    Transmembrane  147 - 163 ( 143 - 164)
   INTEGRAL    Likelihood = -2.92    Transmembrane  122 - 138 ( 120 - 138)
   INTEGRAL    Likelihood = -0.90    Transmembrane   40 -  56 (  39 -  56)
```

```
----- Final Results -----
               bacterial membrane --- Certainty=0.6243(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAD22040 GB:AF118229 transmembrane protein PstA [Streptococcus pneumoniae]
 Identities = 135/263 (51%), Positives = 203/263 (76%)

Query: 23  FFLFAIVYLGAILSFATIAFVVIYILVKGLPHVNTGLFAWTYNTQNVSLLPAFINTIFII 82
            + L  +VY  + L+F ++  ++ +IL+KGLPH++  LF+WTY ++N+SL+PA I+T+ ++
Sbjct: 4   YLLKLLVYCFSALTFGSLFLIIGFILIKGLPHLSLSLFSWTYTSENISLMPAIISTVILV 63

Query: 83  ALTLLFAVPLGIGGSIYLTEYARRDNPYLKIIRVATETLAGIPSIIYGLFGALFFVKYTH 142
               LL A+P+GI    YL EY ++D+  +KI+R+A++TL+GIPSI++GLFG LFFV +
Sbjct: 64  FGALLLALPIGIFAGFYLVEYTKKDSLCVKIMRLASDTLSGIPSIVFGLFGMLFFVVFLG 123

Query: 143 LGLSLISGSLTLSIMILPLIMRTTEEALLSVPDSYREGAFALGAGKLRTIFKIVLPSAMS 202
               SL+SG LT  IM+LP+I+R+TEEALLSV DS R+ ++ LGAGKLRT+F+IVLP AM
Sbjct: 124 FQYSLLSGILTSVIMVLPVIIRSTEEALLSVSDSMRQASYGLGAGKLRTVFRIVLPVAMP 183

Query: 203 GIFAGIILAVGRIIGESAALIFTAGTVAKVAHSVFSSSRTLAVHMYAISGEGLYVDQTYA 262
            GI AG+ILA+GRI+GE+AAL++T GT      S+ SS R+LA+HMY +S EGL+V++ YA
Sbjct: 184 GILAGVILAIGRIVGETAALMYTLGTSTNTPSSLMSSGRSLALHMYMLSSEGLHVNEAYA 243

Query: 263 TAVILLLLVIIVNFVSGLVAKRL 285
            T VIL++ V+++N +S L++++L
Sbjct: 244 TGVILIITVLMINTLSSLLSRKL 266
```

There is also homology to SEQ ID 1686.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 528

A DNA sequence (GBSx0566) was identified in *S.agalactiae* <SEQ ID 1687> which encodes the amino acid sequence <SEQ ID 1688>. Analysis of this protein sequence reveals the following:

```
Possible site: 39
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2687(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 529

A DNA sequence (GBSx0567) was identified in *S.agalactiae* <SEQ ID 1689> which encodes the amino acid sequence <SEQ ID 1690>. This protein is predicted to be transmembrane protein PstC (pstC-2). Analysis of this protein sequence reveals the following:

```
Possible site: 23
>>> Seems to have a cleavable N-term signal seq.
```

```
    INTEGRAL    Likelihood =-10.67   Transmembrane  256 - 272 ( 251 - 279)
    INTEGRAL    Likelihood = -8.86   Transmembrane  141 - 157 ( 133 - 162)
    INTEGRAL    Likelihood = -4.99   Transmembrane  111 - 127 ( 109 - 132)
    INTEGRAL    Likelihood = -4.30   Transmembrane   76 -  92 (  72 -  95)
    INTEGRAL    Likelihood = -1.86   Transmembrane   25 -  41 (  24 -  42)
    INTEGRAL    Likelihood = -1.33   Transmembrane   59 -  75 (  59 -  75)
    INTEGRAL    Likelihood = -0.27   Transmembrane  203 - 219 ( 202 - 219)

----- Final Results -----
              bacterial membrane --- Certainty=0.5267(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAD22039 GB:AF118229 transmembrane protein PstC [Streptococcus pneumoniae]
 Identities = 162/266 (60%), Positives = 212/266 (78%), Gaps = 3/266 (1%)

Query: 15   ITACVSVISAILICLFLFSSGLPAITKIGWGNFIFGKVWHPSN--NIFGIFPMIVGSLYV 72
            ++A V+V++ +LIC F+FS+GLP I   G+   F+ G  W P+N    +GI PMIVGSL +
Sbjct: 1    MSATVAVVAILLICFFIFSNGLPFIANYGFARFLLGSDWSPTNIPASYGILPMIVGSLLI 60

Query: 73   TAGALLLGGPIGILTAVFMAYFCPENIYKPLKSAINLMAGIPSVVYGFFGLVVIVPMIRQ 132
            T GA+++G P GILT+VFM Y+CP+ +Y  LKSAINLMA IPS+VYGFFGL ++VP IR
Sbjct: 61   TLGAIVIGVPTGILTSVFMVYYCPKPVYGFLKSAINLMAAIPSIVYGFFGLQLLVPWIRS 120

Query: 133 YIGGFGMGVLAASILLGIMILPTIVSISESSLRAVPESYYEGGIALGASHERSVFFAVLP 192
            ++G  GM VL AS+LLGIMILPTI+S+SES++R VP++YY G +ALGASHERS+F  +LP
Sbjct: 121 FLGN-GMSVLTASLLLGIMILPTIISLSESAIRTVPKTYYSGSLALGASHERSIFSVILP 179

Query: 193 AAKRGILASVVLGIGRAIGETMAVIMVAGNQAVLPQSLTSGVRTLTTNIVMEMGYSSGLH 252
            AA+ GIL++V+LGIGRA+GETMAVI+VAGNQ ++P  L SG RTLTTNIV+EM Y+SG H
Sbjct: 180 AARSGILSAVILGIGRAVGETMAVILVAGNQPIIPSGLFSGTRTLTTNIVLEMAYASGQH 239

Query: 253 RQALIGTAVVLFIFILMINISFSALQ 278
            R+ALI T+ VLF  IL+IN  F+ L+
Sbjct: 240 REALIATSAVLFFLILLINAYFAYLK 265
```

There is also homology to SEQ ID 1692.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 530

A DNA sequence (GBSx0568) was identified in *S.agalactiae* <SEQ ID 1693> which encodes the amino acid sequence <SEQ ID 1694>. This protein is predicted to be probable hemolysin precursor (pstS). Analysis of this protein sequence reveals the following:

```
Possible site: 34
>>> May be a lipoprotein

----- Final Results -----
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAD22038 GB:AF118229 phosphate binding protein PstS
           [Streptococcus pneumoniae]
 Identities = 134/295 (45%), Positives = 185/295 (62%), Gaps = 9/295 (3%)

Query: 1    MKKHKMLSLLAVSGLMGIGILAGCSNDSSSSSK---GTINIVSREEGSGTRGAFIELFGI 57
            MK  KML+L A+ GL G G++A C N S++S +   GTI ++SRE GSGTRGAF E+ GI
```

```
Sbjct: 1   MKFKKMLTLAAI-GLSGFGLVA-CGNQSAASKQSASGTIEVISRENGSGTRGAFTEITGI 58

Query: 58  ESKNKKGEKVDHTSDAATVTNSTSVMLTTVSKDPSAIGYSSLGSLNSSVKVLKIDGKNAT 117
              K+   +K+D+T+  A + NST  +L+ V  + +AIGY SLGSL  SVK L+IDG  A+
Sbjct: 59  LKKDGD-KKIDNTAKTAVIQNSTEGVLSAVQGNANAIGYISLGSLTKSVKALEIDGVKAS 117

Query: 118 VKDIKSGSYKISRPFNIVTKEGKEKEATKDFIDYILSKDGQAVVEKNGYIPL-DNAKAYQ 176
               +  G Y + RPFNIV      K   +DFI +I SK GQ VV  N +I        Y
Sbjct: 118 RDTVLDGEYPLQRPFNIVWSSNLSK-LGQDFISFIHSKQGQQVVTDNKFIEAKTETTEYT 176

Query: 177 AKVSSGKVVIAGSSSVTPVMEKIKEAYHKVNAKVDVEIQQSDSSTGITSAIDGSADIGMA 236
            ++  SGK+ + GS+SV+ +MEK+ EAY K N +V ++I  + SS GIT+  + +ADIGM
Sbjct: 177 SQHLSGKLSVVGSTSVSSLMEKLAEAYKKENPEVTIDITSNGSSAGITAVKEKTADIGMV 236

Query: 237 SRELDKTESSKGVKATVIATDGIAVVVNKKNKVNDLSTKQVKDIFTGKTTSWSDL 291
            SREL   E  K +    IA DGIAVVVN  NK + +S  ++ D+F+GK T+W  +
Sbjct: 237 SREL-TPEEGKSLTHDAIALDGIAVVVNNDNKASQVSMAELADVFSGKLTTWDKI 290
```

There is also homology to SEQ ID 1696.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

A related GBS gene <SEQ ID 8597> and protein <SEQ ID 8598> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: 23   Crend: 4
McG: Discrim Score:      7.91
GvH: Signal Score (-7.5): -3.72
     Possible site: 34
>>> May be a lipoprotein
ALOM program   count: 0 value:   2.44 threshold:  0.0
   PERIPHERAL  Likelihood =  2.44     248
 modified ALOM score:  -0.99

*** Reasoning Step: 3

----- Final Results -----
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

SEQ ID 1694 (GBS24) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 14 (lane 9; MW 33kDa).

GBS24-His was purified as shown in Figure 194, lane 10.

Example 531

A DNA sequence (GBSx0569) was identified in *S.agalactiae* <SEQ ID 1697> which encodes the amino acid sequence <SEQ ID 1698>. Analysis of this protein sequence reveals the following:

```
Possible site: 30
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.1725(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 532

A DNA sequence (GBSx0570) was identified in *S.agalactiae* <SEQ ID 1699> which encodes the amino acid sequence <SEQ ID 1700>. Analysis of this protein sequence reveals the following:

```
Possible site: 58
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2741(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAB05069 GB:AP001511 unknown conserved protein [Bacillus halodurans]
 Identities = 119/250 (47%), Positives = 149/250 (59%), Gaps = 9/250 (3%)

Query: 1   MQQYFVNGE--AGAYVTIEDKDTIKHMFNVMRLTEDDQVVLVFDDAIKRLAKVVDSSAHR 58
           MQ+YFV  E     YVTI   D +KH+  VMR+T  D+  L+  D   R  +     A+
Sbjct: 1   MQRYFVPKEQMTDTYVTITGDD-VKHIIKVMRMTIGDE--LICSDGHGRTVRCEIEKAND 57

Query: 59  FQIL----EELDNNVEMPVQVTIASGFPKGDKLDFVTQKATELGAAAIWGFPADWSVVKW 114
            ++L    E L  N E+P++VTIA   PKGDKLD++ QK TELGA A W F A  S+VKW
Sbjct: 58  SEVLARVIEPLIPNTELPIRVTIAQALPKGDKLDYIVQKGTELGAQAFWPFSASRSIVKW 117

Query: 115 DGKKLAKKEDKLAKIALGAAEQSKRNRLPQVRLFEKKADFQAELAGFDKIFIAYEESAKE 174
           D KK  KK ++L KIA  AAEQS R R+P +      +    E++GF K  +AYEE AKE
Sbjct: 118 DEKKGRKKTERLMKIAKEAAEQSYRERIPSIETPLAFSKLLQEISGFTKTIVAYEEEAKE 177

Query: 175 GELSALAQNLQTVKAGDKLLFIFGPEGGISPKEIAAFEEVGAIKVGLGPRIMRTETAPLY 234
           G L   A  L  +  GD LL I GPEGG + +EI A +  G    GLGPRI+RTETA LY
Sbjct: 178 GRLMTFAACLNELHHGDSLLVIIGPEGGFTTEEIDAIQRAGGAPAGLGPRILRTETASLY 237

Query: 235 ALSVISYSAE 244
           AL+ ISY  E
Sbjct: 238 ALAAISYHFE 247
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1701> which encodes the amino acid sequence <SEQ ID 1702>. Analysis of this protein sequence reveals the following:

```
Possible site: 56
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2274(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 173/245 (70%), Positives = 202/245 (81%)

Query: 1   MQQYFVNGEAGAYVTIEDKDTIKHMFNVMRLTEDDQVVLVFDDAIKRLAKVVDSSAHRFQ 60
           MQQYF+ G+A   VTI DKDTIKHMF VMRL ++ +VVLVFDD +K LAKV +S AH  +
Sbjct: 1   MQQYFIKGKAEKKVTITDKDTIKHMFQVMRLADEAEVVLVFDDGVKYLAKVTNSMAHELE 60

Query: 61  ILEELDNNVEMPVQVTIASGFPKGDKLDFVTQKATELGAAAIWGFPADWSVVKWDGKKLA 120
           I+E L + VE+PV+VTIASGFPKGDKLD + QK TELGA+A+WG+PADWSVVKWDGKKLA
Sbjct: 61  IIEALPDQVELPVKVTIASGFPKGDKLDTIAQKVTELGASALWGYPADWSVVKWDGKKLA 120

Query: 121 KKEDKLAKIALGAAEQSKRNRLPQVRLFEKKADFQAELAGFDKIFIAYEESAKEGELSAL 180
```

```
           KKEDKLAKI LGAAEQSKRNR+P+V LFE KA+F   L+ FD IFIAYEE+AK G+L+ L
Sbjct: 121 KKEDKLAKIVLGAAEQSKRNRVPEVHLFEHKAEFLKSLSSFDHIFIAYEETAKAGQLATL 180

Query: 181 AQNLQTVKAGDKLLFIFGPEGGISPKEIAAFEEVGAIKVGLGPRIMRTETAPLYALSVIS 240
           A+ ++ VK G K+LFIFGPEGGISP EI  FE   AIKVGLGPRIMR ETAPLYALS +S
Sbjct: 181 AREVKEVKPGAKILFIFGPEGGISPTEITQFEAASAIKVGLGPRIMRAETAPLYALSALS 240

Query: 241 YSAEL 245
           Y+ EL
Sbjct: 241 YALEL 245
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 533

A DNA sequence (GBSx0571) was identified in *S.agalactiae* <SEQ ID 1703> which encodes the amino acid sequence <SEQ ID 1704>. Analysis of this protein sequence reveals the following:

```
Possible site: 34
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -2.28    Transmembrane  238 - 254 ( 237 - 254)

----- Final Results -----
               bacterial membrane --- Certainty=0.1914(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAA82791 GB:AB023064 orf35 [Listeria monocytogenes]
 Identities = 138/309 (44%), Positives = 193/309 (61%), Gaps = 5/309 (1%)

Query: 4   WNELTVHVNREAEEAVSNLLIETGSQGVAISDSADYLGQ-EDRFGELYP---EVEQSDMI 59
           W+E+ VH   EA E V+N+L E G+ GV+I D AD+L + ED+FGE+Y    E    D +
Sbjct: 3   WSEVEVHTTNEAVEPVANVLTEFGAAGVSIEDVADFLREREDKFGEIYALRREDYPEDGV 62

Query: 60  AITAYYPDTLDIEAVKADLADRLANFEGFGLATGSVNLDSQELVEEDWADNWKKYYEPAR 119
            I AY+  T +      ++   L N   F + G       ++ +E+WA  WKKYY P +
Sbjct: 63  IIKAYFLKTTEFVEQIPEIEQTLKNLSTFDIPLGKFQFVVNDVDDEEWATAWKKYYHPVQ 122

Query: 120 ITHDLTIVPSWTDYEAKAGEKIIKMDPGMAFGTGTHPTTKMSLFALEQVLRGGETVIDVG 179
           IT  +TIVPSW  Y   A E II++DPGMAFGTGTHPTT++ + AL   L+ G+ VIDVG
Sbjct: 123 ITDRITIVPSWESYTPSANEIIIELDPGMAFGTGTHPTTQLCIRALSNYLQPGDEVIDVG 182

Query: 180 TGSGVLSIASSLLGAKDIYAYDLDDVAVRVAQENIDMNPGTENIHVAAGDLLKGVQQ-EV 238
           TGSGVLSIAS+ LGAK I A DLD++A R A+ENI +N     I V   +LL+ + +  V
Sbjct: 183 TGSGVLSIASAKLGAKSILATDLDEIATRAAEENITLNKTEHIITVKQNNLLQDINKTNV 242

Query: 239 DVIVANILADILIHLTDDAYRLVKDEGYLIMSGIISEKWDMVRESAEKAGFFLETHMVQG 298
           D++VANILA++++   +D Y+ +K  G  I SGII +K  +V E+ + AG  +E    QG
Sbjct: 243 DIVVANILAEVILLFPEDVYKALKPGGVFIASGIIEDKAKVVEEALKNAGLIIEKMEQQG 302

Query: 299 EWNACVFKK 307
           +W A + K+
Sbjct: 303 DWVAIISKR 311
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1705> which encodes the amino acid sequence <SEQ ID 1706>. Analysis of this protein sequence reveals the following:

```
     Possible site: 34
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -4.57    Transmembrane  238 - 254 ( 237 - 257)

----- Final Results -----
```

```
 bacterial membrane --- Certainty=0.2826(Affirmative) < succ>
  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:BAA82791 GB:AB023064 orf35 [Listeria monocytogenes]
 Identities = 139/309 (44%), Positives = 203/309 (64%), Gaps = 5/309 (1%)

Query: 4   WQEVTVHVHRDAQEAVSHVLIETGSQGVAIADSADYIGQK-DRFGELYP---DVEQSDMI 59
            W EV VH   +A E V++VL E G+ GV+I D AD++ ++ D+FGE+Y    +    D +
Sbjct: 3   WSEVEVHTTNEAVEPVANVLTEFGAAGVSIEDVADFLREREDKFGEIYALRREDYPEDGV 62

Query: 60  AITAYYPSSTNLADIIATINEQLAELASFGLQVGQVTVDSQELAEEDWADNWKKYYEPAR 119
             I AY+  +T    + I  I + L  L++F + +G+      ++ +E+WA  WKKYY P +
Sbjct: 63  IIKAYFLKTTEFVEQIPEIEQTLKNLSTFDIPLGKFQFVVNDVDDEEWATAWKKYYHPVQ 122

Query: 120 ITHDLTIVPSWTDYDASAGEKVIKLDPGMAFGTGTHPTTKMSLFALEQILRGGETVIDVG 179
            IT  +TIVPSW  Y  SA E +I+LDPGMAFGTGTHPTT++ + AL   L+ G+ VIDVG
Sbjct: 123 ITDRITIVPSWESYTPSANEIIIELDPGMAFGTGTHPTTQLCIRALSNYLQPGDEVIDVG 182

Query: 180 TGSGVLSIASSLLGAKTIYAYDLDDVAVRVAQDNIDLNQGTDNIHVAAGDLLKGVSQ-EA 238
            TGSGVLSIAS+ LGAK+I A DLD++A R A++NI LN+    I V   +LL+ +++
Sbjct: 183 TGSGVLSIASAKLGAKSILATDLDEIATRAAEENITLNKTEHIITVKQNNLLQDINKTNV 242

Query: 239 DVIVANILADILVLLTDDAYRLVKKEGYLILSGIISEKLDMVLEAAFSAGFFLETHMVQG 298
            D++VANILA++++L  +D Y+ +K  G  I SGII +K  +V EA  +AG  +E    QG
Sbjct: 243 DIVVANILAEVILLFPEDVYKALKPGGVFIASGIIEDKAKVVEEALKNAGLIIEKMEQQG 302

Query: 299 EWNALVFKK 307
            +W A++ K+
Sbjct: 303 DWVAIISKR 311
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 259/317 (81%), Positives = 287/317 (89%)

Query: 1   MNTWNELTVHVNREAEEAVSNLLIETGSQGVAISDSADYLGQEDRFGELYPEVEQSDMIA 60
            M TW E+TVHV+R+A+EAVS++LIETGSQGVAI+DSADY+GQ+DRFGELYP+VEQSDMIA
Sbjct: 1   METWQEVTVHVHRDAQEAVSHVLIETGSQGVAIADSADYIGQKDRFGELYPDVEQSDMIA 60

Query: 61  ITAYYPDTLDIEAVKADLADRLANFEGFGLATGSVNLDSQELVEEDWADNWKKYYEPARI 120
            ITAYYP + ++  + A + ++LA    FGL  G V +DSQEL EEDWADNWKKYYEPARI
Sbjct: 61  ITAYYPSSTNLADIIATINEQLAELASFGLQVGQVTVDSQELAEEDWADNWKKYYEPARI 120

Query: 121 THDLTIVPSWTDYEAKAGEKIIKMDPGMAFGTGTHPTTKMSLFALEQVLRGGETVIDVGT 180
            THDLTIVPSWTDY+A AGEK+IK+DPGMAFGTGTHPTTKMSLFALEQ+LRGGETVIDVGT
Sbjct: 121 THDLTIVPSWTDYDASAGEKVIKLDPGMAFGTGTHPTTKMSLFALEQILRGGETVIDVGT 180

Query: 181 GSGVLSIASSLLGAKDIYAYDLDDVAVRVAQENIDMNPGTENIHVAAGDLLKGVQQEVDV 240
            GSGVLSIASSLLGAK IYAYDLDDVAVRVAQ+NID+N GT+NIHVAAGDLLKGV QE DV
Sbjct: 181 GSGVLSIASSLLGAKTIYAYDLDDVAVRVAQDNIDLNQGTDNIHVAAGDLLKGVSQEADV 240

Query: 241 IVANILADILIHLTDDAYRLVKDEGYLIMSGIISEKWDMVRESAEKAGFFLETHMVQGEW 300
            IVANILADIL+ LTDDAYRLVK EGYLI+SGIISEK DMV E+A  AGFFLETHMVQGEW
Sbjct: 241 IVANILADILVLLTDDAYRLVKKEGYLILSGIISEKLDMVLEAAFSAGFFLETHMVQGEW 300

Query: 301 NACVFKKTDDISGVIGG 317
            NA VFKKTDDISGVIGG
Sbjct: 301 NALVFKKTDDISGVIGG 317
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 534

A DNA sequence (GBSx0572) was identified in *S.agalactiae* <SEQ ID 1707> which encodes the amino acid sequence <SEQ ID 1708>. Analysis of this protein sequence reveals the following:

```
Possible site: 61
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4198(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 535

A DNA sequence (GBSx0573) was identified in *S.agalactiae* <SEQ ID 1709> which encodes the amino acid sequence <SEQ ID 1710>. This protein is predicted to be transcriptional activator tipa. Analysis of this protein sequence reveals the following:

```
Possible site: 33
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0683(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB15677 GB:Z99122 transcriptional regulator [Bacillus subtilis]
 Identities = 87/246 (35%), Positives = 139/246 (56%), Gaps = 13/246 (5%)

Query: 4   VKEVSILSGVSVRTLHHYDKIGLFPPTALSEAGYRLYDDEALIRLQEILLFRELEFPLKD 63
           VK+V+ +SGVS+RTLHHYD I L  P+AL++AGYRLY D  L RLQ+IL F+E+ F L +
Sbjct: 5   VKQVAEISGVSIRTLHHYDNIELLNPSALTDAGYRLYSDADLERLQQILFFKEIGFRLDE 64

Query: 64  IKYLLEQAKEERQDLLAQQIKLLEWKRSHLEQVITHAKR--LQEKGDDYMN----FDVYN 117
           IK +L+    +R+  L  Q ++L  K+  ++++I    R  L   G + MN    F   +
Sbjct: 65  IKEMLDHPNFDRKAALQSQKEILMKKKQRMDEMIQTIDRTLLSVDGGETMNKRDLFAGLS 124

Query: 118 KTELEQLQA----EAKEKWGQTAA--YKEFAQKHASDDFAQISQEMAKIMVQFGQLKTQN 171
             ++E+ Q     E ++ +G+  A   ++    +++DD+  I  E   I  +
Sbjct: 125 MKDIEEHQQTYADEVRKLYGKEIAEETEKRTSAYSADDWRTIMAEFDSIYRRIAARMKHG 184

Query: 172 VSDESVQMCVKRLQDYISQNFYTCTNEILAGLGQMYQSDDRFSQSIDKAGGAGTSEFVSQ 231
             D  +Q  V   +D+I Q  Y CT +I  GLG++Y +D+RF+ SI++  G G + F+ +
Sbjct: 185 PDDAEIQAAVGAFRDHICQYHYDCTLDIFRGLGEVYITDERFTDSINQY-GEGLAAFLRE 243

Query: 232 AIAYYC 237
           AI  YC
Sbjct: 244 AIIIYC 249
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1711> which encodes the amino acid sequence <SEQ ID 1712>. Analysis of this protein sequence reveals the following:

```
Possible site: 48
```

```
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -8.28   Transmembrane  146 - 162 ( 143 - 167)
   INTEGRAL    Likelihood = -2.92   Transmembrane  172 - 188 ( 171 - 190)

----- Final Results -----
               bacterial membrane --- Certainty=0.4312(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:CAB15677 GB:Z99122 transcriptional regulator [Bacillus subtilis]
 Identities = 40/107 (37%), Positives = 69/107 (64%), Gaps = 6/107 (5%)

Query: 7   YSTGELANLAGVSIRTVQYYDQRGILIPTALTAGGRRLYTDSDLEQLRMICFLRDLGFSI 66
           Y   ++A ++GVSIRT+ +YD   +L P+ALT  G RLY+D+DLE+L+ I F +++GF +
Sbjct: 3   YQVKQVAEISGVSIRTLHHYDNIELLNPSALTDAGYRLYSDADLERLQQILFFKEIGFRL 62

Query: 67  EQIRKVLAEENAAQVLELLLVDHIATAKEDLAAKEQQVDIAVKILDR 113
           ++I+++L   N  +   L        + KE L  K+Q++D  ++ +DR
Sbjct: 63  DEIKEMLDHPNFDRKAAL------QSQKEILMKKKQRMDEMIQTIDR 103
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 40/133 (30%), Positives = 71/133 (53%), Gaps = 6/133 (4%)

Query: 6   EVSILSGVSVRTLHHYDKIGLFPPTALSEAGYRLYDDEALIRLQEILLFRELEFPLKDIK 65
           E++ L+GVS+RT+ +YD+ G+  PTAL+  G RLY D  L +L+ I   R+L F ++ I+
Sbjct: 11  ELANLAGVSIRTVQYYDQRGILIPTALTAGGRRLYTDSDLEQLRMICFLRDLGFSIEQIR 70

Query: 66  YLL--EQAKEERQDLLAQQIKL----LEWKRSHLEQVITHAKRLQEKGDDYMNFDVYNKT 119
            +L  E A +  + LL   I      L  K   ++   +   RL+++    ++F +
Sbjct: 71  KVLAEENAAQVLELLLVDHIATAKEDLAAKEQQVDIAVKILDRLRKQDPQSLDFLMDISL 130

Query: 120 ELEQLQAEAKEKW 132
            ++  +A  K +W
Sbjct: 131 SMKNQKAWKKLQW 143
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 536

A DNA sequence (GBSx0575) was identified in *S.agalactiae* <SEQ ID 1713> which encodes the amino acid sequence <SEQ ID 1714>. Analysis of this protein sequence reveals the following:

```
Possible site: 24
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.06   Transmembrane   57 -  73 (  57 -  73)

----- Final Results -----
               bacterial membrane --- Certainty=0.1022(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB14586 GB:Z99117 yrkN [Bacillus subtilis]
 Identities = 38/136 (27%), Positives = 60/136 (43%), Gaps = 3/136 (2%)

Query: 2   ITLQKAEASDLEKIIA-IQRASFKAVYEKYHDQYDPYVEEVEQIRWKLVERPDCFYHFVL 60
           + L+ A+ SDL +    +Q A   AV E + D  D  +     ++ +    P    + +L
Sbjct: 9   VILELAKESDLPEFQKKLQEAFAIAVIETFGDCEDGPIPSDNDVQ-ESFNAPGAVVYHIL 67

Query: 61  VDETIVGFLRLVIKDEEKRAWLGTAAILPQYQGQGYGSAAMALLEKTYPKLTKWDLCTIA 120
```

```
            D   VG   + I  +     L    + P+Y  QG G +A   +E  YP    W+  T
Sbjct: 68  QDGKNVGGAVVRINSQTNHNSLDLFYVSPEYHSQGIGLSAWKAIEAQYPDTVLWETVTPY 127

Query: 121 QEKLMVSFY-EKCGYH 135
            EK  ++FY  KCG+H
Sbjct: 128 FEKRNINFYVNKCGFH 143
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 537

A DNA sequence (GBSx0576) was identified in *S.agalactiae* <SEQ ID 1715> which encodes the amino acid sequence <SEQ ID 1716>. This protein is predicted to be Bacterial mutT protein. Analysis of this protein sequence reveals the following:

```
Possible site: 13
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2417(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAG06568 GB:AE004742 hypothetical protein [Pseudomonas aeruginosa]
 Identities = 57/131 (43%), Positives = 82/131 (62%)

Query: 10   FSGAKIALFCEGKILTSLRDDFPDLPYAGFWDLPGGGREDNETPLECLFREVDEELSLTL 69
            FSGAK+ALF    ++   RD+ P +P+ G+WD PGGGRE  ETP EC  RE++EE S+ L
Sbjct: 7    FSGAKLALFYGDHLVVYKRDEKPGIPFPGYWDFPGGGREGLETPAECALRELEEEFSIRL 66

Query: 70   TRNHIDWVKTYRGMLKPDKLSVFMVGHISQKEYDSIVLGDEGQDYKLMSIDEFLSHKKVI 129
                I+W + Y         + F+V  +  +E+++I  GDEGQ ++LM +D +L+H   +
Sbjct: 67   EEPRIEWQRQYPSTSGSAPFAYFLVARLEDREFEAIRFGDEGQYWRLMEVDAYLAHAMAV 126

Query: 130 PQLQERLRDYL 140
            P LQ RL DYL
Sbjct: 127 PYLQSRLGDYL 137
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 538

A DNA sequence (GBSx0577) was identified in *S.agalactiae* <SEQ ID 1717> which encodes the amino acid sequence <SEQ ID 1718>. Analysis of this protein sequence reveals the following:

```
Possible site: 46
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3299(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1719> which encodes the amino acid sequence <SEQ ID 1720>. Analysis of this protein sequence reveals the following:

```
Possible site: 41
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.5527(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 111/156 (71%), Positives = 128/156 (81%)

Query: 1    MAKFGFLSVLEEELDKHLQYDFAMDWDKKNHTVEVTFILEAQNSSAIETVDDQGETSSED 60
            MA +GFLSVLEEE+DKH QYD+AMDWDKKNH VEVTF+LEAQN  AI+T+DD GE + +D
Sbjct: 1    MATYGFLSVLEEEMDKHFQYDYAMDWDKKNHAVEVTFVLEAQNKEAIKTIDDSGEVTQDD 60

Query: 61   IVFEDYVLFYNPVKSRFDAEDYLVTIPYEPKKGLSREFLAYFAETLNEVATEGLSDLMDF 120
            IVFEDYVLFYNP KS+FDA DYLVTIP++ KKG SREFLAYFA+ LN+VA EG SDLMDF
Sbjct: 61   IVFEDYVLFYNPAKSQFDAADYLVTIPFDAKKGFSREFLAYFAQFLNDVAIEGHSDLMDF 120

Query: 121 LTDDSIEEFGLSWDTDAFENGRAELKETEFYPYPRY 156
            L DDS  +F L W+  AFE G+  L+E   YPYPRY
Sbjct: 121 LADDSKADFFLEWNAQAFEEGQQGLEEAASYPYPRY 156
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 539

A DNA sequence (GBSx0578) was identified in *S.agalactiae* <SEQ ID 1721> which encodes the amino acid sequence <SEQ ID 1722>. Analysis of this protein sequence reveals the following:

```
Possible site: 26
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2846(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB51273 GB:AL096872 putative acetyltransferase [Streptomyces
          coelicolor A3(2)]
 Identities = 35/109 (32%), Positives = 62/109 (56%), Gaps = 1/109 (0%)

Query: 51   VAEVDDKIAGVLDFGPYYPFPAGKHVATF-GILIAEPYQGQGLGKALLKALLTEAKAQGY 109
            VAE+D   + G +  G    P  +  HV    G+ +A    +G G+G+AL++A + EA+ +G+
Sbjct: 56   VAELDGAVVGYVRLGFPTPLASNTHVRQIRGLAVAGAARGHGVGRALVRAAVEEARHEGF 115

Query: 110  IKIAMHVMGNNSRAISLYQKYGFTEEARITKAFFIENHYVDALIFAKDL 158
             +I + V+G+N+ A  LY+  GF  E    + F ++  YVD ++  + L
Sbjct: 116  RRITLRVLGHNTAARGLYESEGFVVEGVQPEEFHLDGRYVDDVLMGQML 164
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1723> which encodes the amino acid sequence <SEQ ID 1724>. Analysis of this protein sequence reveals the following:

```
Possible site: 18
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
```

```
      bacterial cytoplasm --- Certainty=0.0229(Affirmative) < succ>
       bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
        bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 34/108 (31%), Positives = 59/108 (54%), Gaps = 7/108 (6%)

Query: 35  TESDLEKNLANGMSFFV-----AEVDDKIAGVLDFGPYYPFPAGKHVATFGILIAEPYQG 89
           T  +L   L+   + F+     A +D+K+ G+L+         G+  A   +L+A+ Y+G
Sbjct: 43  TPQELSDFLSRSQTSFIDFCLLARLDEKVVGLLNLSGEV-LSQGQAEADVFMLVAKTYRG 101

Query: 90  QGLGKALLKALLTEAKAQGYIK-IAMHVMGNNSRAISLYQKYGFTEEA 136
            G+G+ LL+  L  A+   YI+ + + V   N++AI LY+KYGF  E+
Sbjct: 102 YGIGQLLLEIALDWAEENPYIESLKLDVQVRNTKAIYLYKKYGFRIES 149
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 540

A DNA sequence (GBSx0579) was identified in *S.agalactiae* <SEQ ID 1725> which encodes the amino acid sequence <SEQ ID 1726>. Analysis of this protein sequence reveals the following:

```
Possible site: 46
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2056(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB14712 GB:Z99118 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 248/417 (59%), Positives = 314/417 (74%), Gaps = 4/417 (0%)

Query: 5   LALRMRPRNINEVIGQQHLVGNGKIIDRMVAANMLSSMILYGPPGIGKTSIASAIAGTTK 64
           LA RMRP  I ++IGQQHLV   KII RMV A  LSSMILYGPPGIGKTSIA+AIAG+T
Sbjct: 4   LAYRMRPTKIEDIIGQQHLVAEDKIIGRMVQAKHLSSMILYGPPGIGKTSIATAIAGSTS 63

Query: 65  YAFRTFNATVDSKKRLQEIAEEAKFSGGLVLLLDEIHRLDKTKQDFLLPLLENGNIIMIG 124
            AFR  NA +++KK ++ +A+EAK SG ++L+LDE+HRLDK KQDFLLP LENG II+IG
Sbjct: 64  IAFRKLNAVINNKKDMEIVAQEAKMSGQVILILDEVHRLDKGKQDFLLPYLENGMIILIG 123

Query: 125 ATTENPFFSVTPAIRSRVQIFELEPLSNEDIKKAIQLAISDKERGF-PFLVTIDDEALDF 183
           ATT NP+ ++ PAIRSR QIFELEPL+ E IK+A++ A+ D+ RG   + V+IDD+A++
Sbjct: 124 ATTANPYHAINPAIRSRTQIFELEPLTPELIKQALERALHDEHRGLGTYSVSIDDQAMEH 183

Query: 184 IVTATNGDLRSAYNSLDLAVMSTSPNEDGSRHISLETMENSLQCSYITMDKNGDGHYDIL 243
                 GD+RSA N+L+LAV+ST  + DG  HI+LET E  LQ    + DK+GD HYD+L
Sbjct: 184 FAHGCGGDVRSALNALELAVLSTKESADGEIHITLETAEECLQKKSFSHDKDGDAHYDVL 243

Query: 244 SALQKSIRGSDVNASLHYAARLVEAGDLPSLARRLTIIAYEDIGLANPEAQIHTVTALEA 303
           SA QKSIRGSD NA+LHY ARL+EAGDL S+ARRL +IAYEDIGLA+P+A    + A++
Sbjct: 244 SAFQKSIRGSDANAALHYLARLIEAGDLESIARRLLVIAYEDIGLASPQAGPRVLNAIQT 303

Query: 304 AQRIGFPEARILIANIVVDLALSPKSNSAYLAMDAALADLRRSGNLPIPRHLRDGHYSGS 363
           A+R+GFPEARI +AN V++L LSPKSNSA LA+D ALAD+R      +P+HL+D HY G+
Sbjct: 304 AERVGFPEARIPLANAVIELCLSPKSNSAILAIDEALADIRAGKIGDVPKHLKDAHYKGA 363

Query: 364 KTLGNARDYKYPHAYPEKWVKQQYLPDKLVGHNYFEANETGKYERALGSNKERIDKL 420
           + LG   DYKYPH Y   WV+QQYLPD L    Y++  +TGK+E AL   K+  DKL
Sbjct: 364 QELGRGIDYKYPHNYDNGWVEQQYLPDPLKNKQYYKPKQTGKFESAL---KQVYDKL 417
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1727> which encodes the amino acid sequence <SEQ ID 1728>. Analysis of this protein sequence reveals the following:

```
Possible site: 60
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2374(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 394/422 (93%), Positives = 409/422 (96%)

Query: 1   MADNLALRMRPRNINEVIGQQHLVGNGKIIDRMVAANMLSSMILYGPPGIGKTSIASAIA 60
           M D+LALRMRP+ I+EVIGQ+HLVG GKII RMV AN LSSMILYGPPGIGKTSIASAIA
Sbjct: 1   MPDHLALRMRPKTISEVIGQKHLVGEGKIIRRMVEANRLSSMILYGPPGIGKTSIASAIA 60

Query: 61  GTTKYAFRTFNATVDSKKRLQEIAEEAKFSGGLVLLLDEIHRLDKTKQDFLLPLLENGNI 120
           GTT+YAFRTFNAT+DSKKRLQEIAEEAKFSGGLVLLLDEIHRLDKTKQDFLLPLLENG I
Sbjct: 61  GTTRYAFRTFNATIDSKKRLQEIAEEAKFSGGLVLLLDEIHRLDKTKQDFLLPLLENGTI 120

Query: 121 IMIGATTENPFFSVTPAIRSRVQIFELEPLSNEDIKKAIQLAISDKERGFPFLVTIDDEA 180
           IMIGATTENPFFSVTPAIRSRVQIFELEPLSNEDIK AIQLAISDKERGFPFLVTIDDEA
Sbjct: 121 IMIGATTENPFFSVTPAIRSRVQIFELEPLSNEDIKTAIQLAISDKERGFPFLVTIDDEA 180

Query: 181 LDFIVTATNGDLRSAYNSLDLAVMSTSPNEDGSRHISLETMENSLQCSYITMDKNGDGHY 240
           LDFIVTATNGDLRSAYNSLDLAVMSTSPNEDGSRHISLETMENSLQ SYITMDKNGDGHY
Sbjct: 181 LDFIVTATNGDLRSAYNSLDLAVMSTSPNEDGSRHISLETMENSLQRSYITMDKNGDGHY 240

Query: 241 DILSALQKSIRGSDVNASLHYAARLVEAGDLPSLARRLTIIAYEDIGLANPEAQIHTVTA 300
           D+LSALQKSIRGSDVNASLHYAARLVEAGDLPSLARRLTIIAYEDIGLANP+AQ+HTVTA
Sbjct: 241 DVLSALQKSIRGSDVNASLHYAARLVEAGDLPSLARRLTIIAYEDIGLANPDAQVHTVTA 300

Query: 301 LEAAQRIGFPEARILIANIVVDLALSPKSNSAYLAMDAALADLRRSGNLPIPRHLRDGHY 360
           L+AAQRIGFPEARI IAN+V+DLALSPKSNSAYLAMDAALADLR SGNLPIPRHLRDGHY
Sbjct: 301 LDAAQRIGFPEARIPIANVVIDLALSPKSNSAYLAMDAALADLRTSGNLPIPRHLRDGHY 360

Query: 361 SGSKTLGNARDYKYPHAYPEKWVKQQYLPDKLVGHNYFEANETGKYERALGSNKERIDKL 420
           +GSK LGNA+DY YPHAYPEKWVKQQYLPDKLVGH+YFEANETGKYERALGSNKERIDKL
Sbjct: 361 AGSKDLGNAKDYLYPHAYPEKWVKQQYLPDKLVGHHYFEANETGKYERALGSNKERIDKL 420

Query: 421 SD 422
           SD
Sbjct: 421 SD 422
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 541

A DNA sequence (GBSx0580) was identified in *S.agalactiae* <SEQ ID 1729> which encodes the amino acid sequence <SEQ ID 1730>. Analysis of this protein sequence reveals the following:

```
Possible site: 16
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2991(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10207> which encodes amino acid sequence <SEQ ID 10208> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 542

A DNA sequence (GBSx0581) was identified in *S.agalactiae* <SEQ ID 1731> which encodes the amino acid sequence <SEQ ID 1732>. Analysis of this protein sequence reveals the following:

```
Possible site: 29
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2402(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 543

A DNA sequence (GBSx0582) was identified in *S.agalactiae* <SEQ ID 1733> which encodes the amino acid sequence <SEQ ID 1734>. Analysis of this protein sequence reveals the following:

```
Possible site: 49
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-10.40   Transmembrane  231 - 247 ( 225 - 250)
   INTEGRAL    Likelihood = -9.92   Transmembrane  159 - 175 ( 151 - 179)
   INTEGRAL    Likelihood = -9.08   Transmembrane   21 -  37 (  18 -  43)
   INTEGRAL    Likelihood = -9.08   Transmembrane  181 - 197 ( 176 - 201)
   INTEGRAL    Likelihood = -3.35   Transmembrane  111 - 127 ( 110 - 130)
   INTEGRAL    Likelihood = -2.81   Transmembrane   74 -  90 (  74 -  93)

----- Final Results -----
               bacterial membrane --- Certainty=0.5161(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB15891 GB:Z99123 yxlG [Bacillus subtilis]
 Identities = 54/203 (26%), Positives = 100/203 (48%), Gaps = 7/203 (3%)

Query: 1    MTGLIPMLKKEWLENSRSHKALALLLISIIFGILGPLTALLMPEIMA--GILPKKLQEAI 58
            M  ++ +L+KEWLE  +S K + L +  +I G+  PLT   MPEI+A  G LP  ++ +
Sbjct: 1    MKVMMALLQKEWLEGWKSGKLIWLPIAMMIVGLTQPLTIYYMPEIIAHGGNLPDGMKISF 60

Query: 59   PDPTYLDSYSQYFKNINQLGLILLVFLFSGSLTQEFTRGTLINLITKGLSKKAIILAKFI 118
              P+  +        N LG+ L++F   GS+  E  +G    ++++ ++    I++K++
Sbjct: 61   TMPSGSEVMVSTLSQFNTLGMALVIFSVMGSVANERNQGVTALIMSRPVTAAHYIVSKWL 120
```

```
Query: 119 MMTLIWSISYILGSLTQYAYTLYYFNNHGQHKLIV-YGTSWIFGLLLLSLILFYSVIFRK 177
           + ++I  +S+  G    Y Y    F +    +     G   ++ + +++  L  S IFR
Sbjct: 121 IQSVIGIMSFAAGYGLAYYYVRLLFEDASFSRFAASLGLYALWVIFIVTAGLAGSTIFR- 179

Query: 178 TAGVLIAC---LMTIVAFFISGF 197
           + G   AC   L   V+F +  F
Sbjct: 180 SVGAAAACGIGLTAAVSFAVHYF 202
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 544

A DNA sequence (GBSx0583) was identified in *S.agalactiae* <SEQ ID 1735> which encodes the amino acid sequence <SEQ ID 1736>. This protein is predicted to be ABC transporter, ATP-binding protein. Analysis of this protein sequence reveals the following:

```
Possible site: 61
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.1344(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB15892 GB:Z99123 similar to ABC transporter (ATP-binding
           protein) [Bacillus subtilis]
 Identities = 116/303 (38%), Positives = 175/303 (57%), Gaps = 18/303 (5%)

Query: 4   ISLQNLSKSFGDQIILNQVSLELEENKIYGFVGPNGAGKTTTIKMILGLLKVDSGTISVM 63
           +S+++L KS+     +  VS  + EN+    +GPNGAGKTTT++M+ GLL   SGTI ++
Sbjct: 2   LSIESLCKSYRHHEAVKNVSFHVNENECVALLGPNGAGKTTTLQMLAGLLSPTSGTIKLL 61

Query: 64  GNPVTFGQTKSNQVIGYLPDVPEFYDYMTAQEYLQLC---AGLAQNKTSLPIADLLEQVG 120
           G      +    ++IGYLP  P FY +MTA E+L      +GL++ K    I ++LE VG
Sbjct: 62  GE-----KKLDRRLIGYLPQYPAFYSWMTANEFLTFAGRLSGLSKRKCQEKIGEMLEFVG 116

Query: 121 LADN-QQRISTYSRGMKQRLGLAQALIHNPKILICDEPTSALDPQGRQEILSIISQLRGQ 179
           L +   +RI  YS GMKQRLGLAQAL+H PK LI DEP SALDP GR E+L ++ +L+
Sbjct: 117 LHEAAHKRIGGYSGGMKQRLGLAQALLHKPKFLILDEPVSALDPTGRFEVLDMMRELKKH 176

Query: 180 KTVIFSTHILSDVEKVCDQVLILTKSGIH---NLEDLRDKASASVNQLNLLIKVSDNEAQ 236
             V+FSTH+L D E+VCDQV+I+    I     L++L+ +   +V  L++  K+     +
Sbjct: 177 MAVLFSTHVLHDAEQVCDQVVIMKNGEISWKGELQELKQQQQTNVFTLSVKEKLEGWLEE 236

Query: 237 KLALRFPLNQKDQYYKVHLELSEANNREQALASFYRYLVEQEITPYFIELLEDSLEDFYL 296
           K  +   + +     +   EL + +     L+      + + +T    E   +SLED YL
Sbjct: 237 KPYVSAIVYKNPS--QAVFELPDIHAGRSLLSD----CIRKGLTVTRFEQKTESLEDVYL 290

Query: 297 EVI 299
           +V+
Sbjct: 291 KVV 293
```

There is also homology to SEQ ID 686.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 545

A DNA sequence (GBSx0584) was identified in *S.agalactiae* <SEQ ID 1737> which encodes the amino acid sequence <SEQ ID 1738>. Analysis of this protein sequence reveals the following:

```
Possible site: 32
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4383(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAB71491 GB:U53767 ORF6 [Bacillus pumilus]
 Identities = 25/60 (41%), Positives = 41/60 (67%)

Query: 2  IGDTILFERTRLGMTQEKLSDYLHLTKATISKWENNQAKPDIDYLILMAKLFDMTLDELV 61
           +G  I  +R  L ++QE +++ L +++  ISKWE NQ++P +D LI +A+LFD  + ELV
Sbjct: 4  LGSNISNKRKSLKLSQEYVAEQLGVSRQAISKWETNQSEPSMDNLIRLAELFDSDIKELV 63
```

There is also homology to SEQ ID 1740.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 546

A DNA sequence (GBSx0585) was identified in *S.agalactiae* <SEQ ID 1741> which encodes the amino acid sequence <SEQ ID 1742>. Analysis of this protein sequence reveals the following:

```
Possible site: 41
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4241(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB15470 GB:Z99121 yvdC [Bacillus subtilis]
 Identities = 59/104 (56%), Positives = 76/104 (72%)

Query: 1   MDITAYQKWVSEFYKKRNWYQYNSFIRSNFLCEEVGELAQAIRKYEIGRDRPDEIEKSNN 60
            M +   +KW+ EFY+KR W +Y  FIR  FL EE GELA+A+R YEIGRDRPDE E S
Sbjct: 1   MQLADAEKWMKEFYEKRGWTEYGPFIRVGFLMEEAGELARAVRAYEIGRDRPDEKESSRA 60

Query: 61  ENLNDIKEELGDVLDNIFILADQYNISLEEIIEAHKNKLEKRFE 104
            E   ++ EE+GDV+ NI ILAD Y +SLE++++AH+ KL KRFE
Sbjct: 61  EQKQELIEEMGDVIGNIAILADMYGVSLEDVMKAHQEKLTKRFE 104
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 547

A DNA sequence (GBSx0586) was identified in *S.agalactiae* <SEQ ID 1743> which encodes the amino acid sequence <SEQ ID 1744>. Analysis of this protein sequence reveals the following:

```
Possible site: 61
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0453(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAB06803 GB:AP001517 unknown conserved protein [Bacillus halodurans]
 Identities = 87/187 (46%), Positives = 125/187 (66%)

Query: 1   MKITVFCGASNGNNPIYSQKIVELGEWMIKNNHDLVYGGGKVGLMGVIADTVINNGGQAI 60
            MKI VFCG+SNG + +Y +   +LG+ + +    LVYGG  VG+MG +AD+V+  GG+ I
Sbjct: 1   MKIAVFCGSSNGASDVYKEGARQLGKELARRGITLVYGGASVGIMGAVADSVLEAGGEVI 60

Query: 61  GVIPTFLKDREIAHTNLSKLIVVENMPQRKGKMMSLGEAYIALPGGPGTLEEISEVISWS 120
            GV+P FL++ EI+H +L+KLIVVE M +RK KM  L + ++ALPGGPGTLEE  E+ +W+
Sbjct: 61  GVMPRFLEEPEISHPHLTKLIVVETMHERKAKMAELADGFLALPGGPGTLEEFFEIFTWA 120

Query: 121 RIGQNDSPCILYNINGYFNHLESMFDHMVSEGFLSQNDRNNVLFSDDIIEIEKFIKDYQS 180
            +IG +  PC L NIN YF+ L ++  HM +E FL +  R+  L   D I +      Y+
Sbjct: 121 QIGLHQKPCGLLNINHYFDPLVTLLHHMSNEQFLHEKYRSMALVHTDPILLLDQFSTYEP 180

Query: 181 PTIRKYS 187
            PT++ YS
Sbjct: 181 PTVKAYS 187
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 548

A DNA sequence (GBSx0587) was identified in *S.agalactiae* <SEQ ID 1745> which encodes the amino acid sequence <SEQ ID 1746>. Analysis of this protein sequence reveals the following:

```
Possible site: 16
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.5288(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 549

A DNA sequence (GBSx0588) was identified in *S.agalactiae* <SEQ ID 1747> which encodes the amino acid sequence <SEQ ID 1748>. This protein is predicted to be integrase. Analysis of this protein sequence reveals the following:

```
Possible site: 38
>>> Seems to have no N-terminal signal sequence
```

```
----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3685(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAF12706 GB:AF066865 integrase [bacteriophage TPW22]
 Identities = 106/377 (28%), Positives = 199/377 (52%), Gaps = 31/377 (8%)

Query: 4   ARYRRRGNQNLWAYEIREEGKTVAYNS----GFKTKKLAEAEAEPILQKLRTGSIITKNI 59
           A +R+RG    W + +  +     Y      G+KTKK AEA A+   ++L   S    +I
Sbjct: 2   ANFRKRGKT--WQFRLSYKDNNGEYKKFEKGGYKTKKEAEAAADEAKKRLNNHSEFDNDI 59

Query: 60  SLPELYQEWLDLKIMPSNRSDVTKKKYLSRKVTLEKLFGDKPISQIRPSEYQRIMNNYGQ 119
           SL + +++W  +   P + ++ T + Y      ++K   DKPI++I P+ YQ ++N
Sbjct: 60  SLYDFFEKWAKVYKKP-HVTEATWRTYKRTLNLIDKYIKDKPIAEITPTFYQAVLNKMSL 118

Query: 120 RVSRNFLGRLNTGVKQSLQMAIADKVMIEDFTQNVELFSTVKSQDADSKYLHSEKAYLDL 179
              +  L +    +K ++++A+ +KV+ E+F    +  S + ++  + KYLH+++ YL L
Sbjct: 119 LYRQESLDKFYFQIKSAMKIAVHEKVISENFADFTKAKSKLAARPVEEKYLHADE-YLKL 177

Query: 180 INAVKDKFNYKKSVVPYIIYFLLKTGMRYGELIALTWEDIDFDKGIFKTYRRFN-SETSQ 238
           +   ++K  Y      +  Y    TGMR+ EL+ LTW  +DFDK      R ++ S T+
Sbjct: 178 LAIAEEKMEYTSY---FACYLTAVTGMRFAELLGLTWSHVDFDKKEISIQRTWDYSITNN 234

Query: 239 FVPPKNKTSIRIVPVDNECLEILKNLKIEQNQSNKELGLQNTNNMVFQHFGYPNSVPSTN 298
           F   KN++S R +P+ ++ +++LK  K       KE   +N  + V  +        S N
Sbjct: 235 FAETKNESSKRKIPISSKTIKLLKKYK-------KEYWHENKYDRVIYNL-------SNN 280

Query: 299 GTNKVLRGIVQELNIEPIITTKGARHTYGSFLWHRGYDLGIIAKILGHKDISMLIEVYGH 358
           G NK ++ ++    + P       RH++ S+L ++G DL  ++K+LGH+++++ ++VY H
Sbjct: 281 GLNKTIK-VIAGRKVHP----HSLRHSFASYLIYKGIDLLTVSKLLGHENLNVTLKVYAH 335

Query: 359 TLEEKIQEEYNEIKQLW 375
            L+E  QE  + I++++
Sbjct: 336 QLKEMEQENNDVIRKIF 352
```

There is also homology to SEQ ID 578.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 550

A DNA sequence (GBSx0589) was identified in *S.agalactiae* <SEQ ID 1749> which encodes the amino acid sequence <SEQ ID 1750>. Analysis of this protein sequence reveals the following:

```
Possible site: 54
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2710(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 551

A DNA sequence (GBSx0590) was identified in *S.agalactiae* <SEQ ID 1751> which encodes the amino acid sequence <SEQ ID 1752>. Analysis of this protein sequence reveals the following:

```
Possible site: 51
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2534(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAA06248 GB:D29979 ORF3 [Bacillus stearothermophilus]
 Identities = 81/263 (30%), Positives = 135/263 (50%), Gaps = 14/263 (5%)

Query: 65  MGVHVELKGQGCRQYEEFIEGNDNNWTSLVKRLI-DNNSNFTRLDIANDIFDESLNVQRL 123
            MG+HVE+ GQGCR +E        NW  L  RL+ +   N TRLD+A D F     +  L
Sbjct: 1   MGIHVEMTGQGCRLFELH---TSINWYELFYRLVYEYEVNITRLDVAVDDFKGYFKINTL 57

Query: 124 YEYSKKGLCITTARHAEYHEKFVIDSGELVGETVVFGARGNQQWCVYNKLMEQNGKLQTD 183
             +  K     +  + A + E  VI+ GE +G T+ FGA  +    +  +  E+N ++  D
Sbjct: 58  VKKLKDDEVTSRFKKARHIENIVIEGGETIGHTLYFGAPSSD---IQVRFYEKNVQMGMD 114

Query: 184 IDINSWVRAELRCWQEKANLIAHQL-NDMRPLASIYFEAINGHYRFVSPKARDKNKRRRE 242
            ID+  W R E++   ++A+++A  + +D+ PL  I    +  + +F + KA DKNK+R
Sbjct: 115 IDV--WNRTEIQLRDDRAHVVAQIIADDVLPLGEIVAGLLRNYIQFRTRKATDKNKKRWP 172

Query: 243 SVRWWQNYINTEEKTRLSIVREKPTLRQSEAWTDKQVSKTIAKVYMAKYEAYGIDQAEVF 302
              R+W N++   +  R++    K ++ +   W D QVSK+   +Y    E    ++ + F
Sbjct: 173 LARFWLNFLGDVQPLRIAKQMPKTSIEKKYRWIDSQVSKSFFMIYYCLNE----EEKQRF 228

Query: 303 LQDLLRRGVEKFTDNDEKEIEQY 325
            + D+L  G  K T  D + I Q+
Sbjct: 229 IDDVLAEGASKLTKADLQVINQF 251
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 552

A DNA sequence (GBSx0591) was identified in *S.agalactiae* <SEQ ID 1753> which encodes the amino acid sequence <SEQ ID 1754>. Analysis of this protein sequence reveals the following:

```
Possible site: 13
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2700(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 553

A DNA sequence (GBSx0592) was identified in *S.agalactiae* <SEQ ID 1755> which encodes the amino acid sequence <SEQ ID 1756>. Analysis of this protein sequence reveals the following:

```
Possible site: 50
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3121(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1757> which encodes the amino acid sequence <SEQ ID 1758>. Analysis of this protein sequence reveals the following:

```
Possible site: 24
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2913(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 19/52 (36%), Positives = 33/52 (62%)

Query: 8  FGPNLTRLRKERGISQVELSNQLQIGKQSISDYEKQKAFPTFANLDKIAEYF 59
           F  NL  L  ++ I Q+++ N+L I K +I+ Y K ++ PT  N+ K+A++F
Sbjct: 15 FSTNLNMLMAKKNIKQIDIHNKLGIPKSTITGYVKGRSLPTAGNVQKLADFF 66
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 554

A DNA sequence (GBSx0593) was identified in *S.agalactiae* <SEQ ID 1759> which encodes the amino acid sequence <SEQ ID 1760>. Analysis of this protein sequence reveals the following:

```
Possible site: 54
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
                bacterial outside --- Certainty=0.3000(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAA98584 GB:L44593 ORF536; putative [Lactococcus phage BK5-T]
 Identities = 248/532 (46%), Positives = 359/532 (66%), Gaps = 16/532 (3%)

Query: 1   MNFIEQISENNQFPIIFVGSGITQRYFENAPTWEKLLKDIWLELFDEESYYAK--AFELR 58
            MNFIE I +NNQFPIIFVGSG+T+RYF+N   WE+LL ++W  + +E+++Y +   FE
Sbjct: 1   MNFIENIKDNNQFPIIFVGSGVTKRYFKNGLKWEQLLLELWNLVEEEKAFYTQYHVFENL 60

Query: 59  ERFEN-----NDFDIYTNLASLLEKEVSKAFINGNIQVDNLDLKTAYELNISPFKQLVAN 113
             + +N      +F+I   +A +LE++++ AF +  + +DNL L  A+  +ISPF+Q +AN
Sbjct: 61  LKSKNLSKSDKEFEINLMMAGILEEKINNAFYSDELNIDNLTLAQAHTEHISPFRQCIAN 120

Query: 114 RFSNLKIREEKIEEIKQFSQMLSKARIIITTNYDNFIEECLKTINVSVKINVGNKGLFLK 173
```

```
            FSNL  ++   EEI  FS+ML KAR I+TTNYDNFIEEC    NVS+K+NVGN GLF+K
Sbjct: 121 TFSNLDRKKGFDEEIISFSKMLVKARFIVTTNYDNFIEECFSKRNVSIKVNVGNSGLFVK 180

Query: 174 SSDYGELYKIHGTVDDASTITITKEDYEKNVTKSALINAKILSNLVESPILFLGYSLTDE 233
           S+DYGELYKIHG+V + +TI IT EDY+ N +K AL+NAKILSNL ESPILF+GYSLTD+
Sbjct: 181 SNDYGELYKIHGSVKNPNTICITSEDYKNNESKLALVNAKILSNLTESPILFIGYSLTDK 240

Query: 234 NIRKLLTDFAENSPFDISESAQKIGVVEYLPDSESIETVVSSLPDLSVYYSCLKTDNFTN 293
           NIR+LLT ++EN P++ISE+A +IGVVEY PD   I+ +VS++PDL ++Y+ + TDN+
Sbjct: 241 NIRELLTSYSENLPYEISEAAARIGVVEYTPDKIEIQDIVSNIPDLGIHYTKISTDNYKK 300

Query: 294 IYRLISKINQGFLPSEIAKYENVFRKIIEVKGESKDLKTVLTSYEDLANLTEDEIRSKNI 353
           IY  IS+I QG+LPSEIAK+E  FRKIIEVKG+ K+L TVLTS+ D++ +  +E+++KNI
Sbjct: 301 IYDEISQIEQGYLPSEIAKFEGAFRKIIEVKGKEKELDTVLTSFIDISKINTEELKNKNI 360

Query: 354 VVAFGDERYIYKFPDFKEYVRSYFLDKETIPQEIVIRFIATQPVASHLPIKKYMFAMSEY 413
           VVAFGD +YIYK P +K+Y+R YF +   +   I + F+  +     +P KK+M  +  +
Sbjct: 361 VVAFGDSKYIYKMPTYKDYIREYFSNSMELDTRIALLFLKKRSANYPVPYKKHMGVIESW 420

Query: 414 --ISKDSNKYTENIKKRLSKEEELSLDDFTSSIGVPLL--HSKTLERQTEIVGILE-ADV 468
             I  D  +  E++K R+S   E  +  ++      L   +   L + + I  ++  ++V
Sbjct: 421 GSIPNDLVQEVESLKTRISNFPESIVRTYSIKANKDLAKKYLPYLNKTSTIEDVMSLSNV 480

Query: 469 PDNVRYNFIATHIKNFPKEELFLLVEKIID----EGIFETSRRRFLKAFDLL 516
           P   +  FI   I  F  EEL   + K ID    +GI  T  R+ + ++ ++
Sbjct: 481 PLYNKLRFILFKIDKFKVEELKDFIVKNIDMGEGKGISSTLYRKIVMSYSII 532
```

A related GBS gene <SEQ ID 8599> and protein <SEQ ID 8600> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 8
McG: Discrim Score:      1.55
GvH: Signal Score (-7.5): 0.27
     Possible site: 54
>>> Seems to have a cleavable N-term signal seq.
ALOM program   count: 0 value:   2.44 threshold:  0.0
   PERIPHERAL  Likelihood =  2.44     214
 modified ALOM score:  -0.99

*** Reasoning Step: 3

----- Final Results -----
                bacterial outside --- Certainty=0.3000(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
47.3/69.1% over 531aa
                                                                 Lactococcus lactis
  EGAD|36707| hypothetical protein Insert characterized
  GP|928833|gb|AAA98584.1||L44593 ORF536; putative {Lactococcus lactis phage BK5-T} Insert
characterized
  PIR|T13261|T13261 hypothetical protein 536 - phage BK5-T Insert characterized

ORF00184(301 - 1848 of 2154)
EGAD|36707|38110(1 - 532 of 536) hypothetical protein {Lactococcus
lactis}GP|928833|gb|AAA98584.1||L44593 ORF536; putative {Lactococcus lactis phage BK5-
T}PIR|T13261|T13261 hypothetical protein 536 - Lactococcus lactis phage BK5-T
%Match = 32.3
%Identity = 47.2  %Similarity = 69.0
Matches = 247  Mismatches = 155  Conservative Sub.s = 114

126       156       186       216       246       276       306       336
RMLILKAFYLAKFLKYYC*KK*CGTKRGQLYFRVYGLIIKINKMVSKML**D*QLNKLIINKR*GQELVNFIEQISENNQ
                                                                    :|||| | :|||
                                                                    MNFIENIKDNNQ
```

```
                                                                            10

366       396       426       456       474       495       525       555
FPIIFVGSGITQRYFENAPTWEKLLKDIWLELFDEESYYAK--AFE--LRER---FENNDFDIYTNLASLLEKEVSKAFI
|||||||||:|:|||:|   ||:|| ::|  : :|:::| :   ||  |: :     : :|:|   :| :||:::: ||
FPIIFVGSGVTKRYFKNGLKWEQLLLELWNLVEEEKAFYTQYHVFENLLKSKNLSKSDKEFEINLMMAGILEEKINNAFY
                30        40        50        60        70        80        90

585       615       645       675       705       735       765       795
NGNIQVDNLDLKTAYELNISPFKQLVANRFSNLKIREEKIEEIKQFSQMLSKARIIITTNYDNFIEECLKTINVSVKINV
:  : :||| |  |:  :||||:| :|| ||||  ::   |||  ||:|| ||| |:|||||||||||:   |||:|:||
SDELNIDNLTLAQAHTEHISPFRQCIANTFSNLDRKKGFDEEIISFSKMLVKARFIVTTNYDNFIEECFSKRNVSIKVNV
               110       120       130       140       150       160       170

825       855       885       915       945       975       1005      1035
GNKGLFLKSSDYGELYKIHGTVDDASTITITKEDYEKNVTKSALINAKILSNLVESPILFLGYSLTDENIRKLLTDFAEN
|| |||:||:||||||||||:| : :|| || |||: | :| ||:|||||||| ||||||:||||||:|||:||| ::||
GNSGLFVKSNDYGELYKIHGSVKNPNTICITSEDYKNNESKLALVNAKILSNLTESPILFIGYSLTDKNIRELLTSYSEN
               190       200       210       220       230       240       250

1065      1095      1125      1155      1185      1215      1245      1275
SPFDISESAQKIGVVEYLPDSESIETVVSSLPDLSVYYSCLKTDNFTNIYRLISKINQGFLPSEIAKYENVFRKIIEVKG
 |::|||:| :|||||| ||   |: :||::||| ::|: : |||:  ||  ||:| ||:|||||||:|  |||||||||
LPYEISEAAARIGVVEYTPDKIEIQDIVSNIPDLGIHYTKISTDNYKKIYDEISQIEQGYLPSEIAKFEGAFRKIIEVKG
               270       280       290       300       310       320       330

1305      1335      1365      1395      1425      1455      1485      1515
ESKDLKTVLTSYEDLANLTEDEIRSKNIVVAFGDERYIYKFPDFKEYVRSYFLDKETIPQEIVIRFIATQPVASHLPIKK
: |:| |||||: |:: :  :|:::||||||||| :|||| | :|:|:| || :   :   | : |:  :     :| ||
KEKELDTVLTSFIDISKINTEELKNKNIVVAFGDSKYIYKMPTYKDYIREYFSNSMELDTRIALLFLKKRSANYPVPYKK
               350       360       370       380       390       400       410

1569      1599      1629                          1683      1710      1740
YMFAMSEY--ISKDSNKYTENIKKRLSKEEELSLDDFTSSIGVPLLHS--KTLERQTEIVGILE-ADVPDNVRYNFIATH
:|  :  :  |  |  :  |::| |:|   |  :  ::      |       | : : |  ::  ::||   :  ||
HMGVIESWGSIPNDLVQEVESLKTRISNFPESIVRTYSIKANKDLAKKYLPYLNKTSTIEDVMSLSNVPLYNKLRFILFK
               430       440       450       460       470       480       490

1770                1818      1848      1878      1908      1938      1968
IKNFPKEELFLLVEKIID----EGIFETSRRRFLKAFDLLHY*IKKSQHCYAMRDFFWTTINKRYENNCFYLTSILQYIF
|  |  |||  :: | ||    :||  |  |: : :: ::   |
IDKFKVEELKDFIVKNIDMGEGKGISSTLYRKIVMSYSIITEGI
               510       520       530
```

No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 8600 (GBS142) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 18 (lane 5; MW 54kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 33 (lane 6; MW 79.8kDa).

The GBS142-GST fusion product was purified (Figure 195, lane 3) and used to immunise mice. The resulting antiserum was used for Western blot (Figure 249). These tests confirm that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 555

A DNA sequence (GBSx0594) was identified in *S.agalactiae* <SEQ ID 1761> which encodes the amino acid sequence <SEQ ID 1762>. This protein is predicted to be integrase. Analysis of this protein sequence reveals the following:

```
Possible site: 37
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.2933(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAA98585 GB:L44593 integrase [Lactococcus phage BK5-T]
 Identities = 124/382 (32%), Positives = 202/382 (52%), Gaps = 21/382 (5%)

Query: 1   MATYRQRGKKKLWDYRIFNEKSELVA-SGSGFKTKREAMNEAMRIE---QQKLLVNSISS 56
           MATY++RGK   W Y I   K  L   +  GF TK +A  EAM IE   ++  +V+ I
Sbjct: 1   MATYQKRGKT--WQYSISRTKQGLPRLTKGGFSTKSDAQAEAMDIESKLKKGFIVDPIKQ 58

Query: 57  DITLYDL-WFEWYSLIIKPSNLAETTKNKYFTRGSVIRKLFGNQKVNKIKHSAYQRKLNT 115
           +I+ Y   W E Y    K + + E T   Y      ++    N  +++I  S+YQR LN
Sbjct: 59  EISEYFKDWMELY----KKNAIDEMTYKGYEQTLKYLKTYMPNVLISEITASSYQRALNK 114

Query: 116 YAEKYTKNHVRRLNSDIKKAIQFAKRDGVLLSDFTDGVVIAGRKFVKDADDKYLHSIFD- 174
           +AE + K   +  ++ ++ +IQ    +G L  DFT   V+ G    K   DK+++  FD
Sbjct: 115 FAETHAKASTKGFHTRVRASIQPLIEEGRLQKDFTTRAVVKGNGNDKAEQDKFVN--FDE 172

Query: 175 YKKVISYLENNLD--YSNSIVYYLLLVLFKTGLRVGEALALTWDDVNFEDLEIKTYR--R 230
           YK+++ Y  N L+  YS+  + +++ +   TG+R  EA  L WDD++F +  IK  R
Sbjct: 173 YKQLVDYFRNRLNPNYSSPTMLFIISI---TGMRASEAFGLVWDDIDFNNNTIKCRRTWN 229

Query: 231 FSGDKGTFSPPKTKTSIRTIPISQSLALILRDLKDDQQVMLKNLKIVNMNNQIFYDYRYG 290
           +    G F  PKT   IR I I      +L+D ++ Q+ + ++L I  +++ + Y
Sbjct: 230 YRNKVGGFKKPKTDAGIRDIVIDDESMQLLKDFREQQKTLFESLGIKPIHDFVCYHPYRK 289

Query: 291 VSTNSAINKSLKNVLKILNINSKMTATGARHTYGSYLLAKGVDIWVVARLMGHKDITQLL 350
           + T SA+  +L + LK LNI++ +T  G RHT+ S LL  GVDI  V++ +GH  +
Sbjct: 290 IITLSALQNTLDHALKKLNISTPLTIHGLRHTHASVLLYHGVDIMTVSKRLGHASVAITQ 349

Query: 351 ETYGHVLTEVINKEYETVRSLV 372
           +TY H++ E+ NK+ + +  L+
Sbjct: 350 QTYIHIIKELENKDKDKIIELL 371
```

There is also homology to SEQ ID 578.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 556

A DNA sequence (GBSx0595) was identified in *S.agalactiae* <SEQ ID 1763> which encodes the amino acid sequence <SEQ ID 1764>. Analysis of this protein sequence reveals the following:

```
Possible site: 54
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.1603(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10209> which encodes amino acid sequence <SEQ ID 10210> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAB07266 GB:AP001519 unknown conserved protein in others
```

```
            [Bacillus halodurans]
 Identities = 26/71 (36%), Positives = 39/71 (54%), Gaps = 6/71 (8%)

Query: 37  WWDIDNLQELLGIGRSKLINDILLNPDIKKEVDLSINPNGFIVYPKGKGSRYKILATK-- 94
            WW + +L+E  G     L  +ILL+P  K  +D  I   GF+ YP+ KG R+  +A+
Sbjct: 4   WWSMQDLKERTGYSEDWLKENILLHPRYKPMLD--IENGGFVYYPEKKGERWCFIASSME 61

Query: 95  --ARKYFEDNF 103
               +KYF+D F
Sbjct: 62  EFLKKYFKDIF 72
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 557

A DNA sequence (GBSx0596) was identified in *S.agalactiae* <SEQ ID 1765> which encodes the amino acid sequence <SEQ ID 1766>. Analysis of this protein sequence reveals the following:

```
Possible site: 14
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -3.88   Transmembrane    12 -  28 (  11 -  29)

----- Final Results -----
               bacterial membrane --- Certainty=0.2550(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAB99663 GB:U67604 chromosome segretation protein (smc1)
           [Methanococcus jannaschii]
 Identities = 53/210 (25%), Positives = 95/210 (45%), Gaps = 33/210 (15%)

Query: 20  IFTNVGVLISNSRDNKAIQRELELLEEGQEKLVDEFSKISTNQYDKYV----------LI 69
            +F  +G+L  N      + + + +    + K++DE S I+     K            LI
Sbjct: 133 LFRRLGLLGDNVISQGDLLKIINISPIERRKIIDEISGIAEFDEKKKKAEEELKKARELI 192

Query: 70  Q------SNLSNNIEKNKQELVQKNSYVK--EDTKYIRDEMLIEKKSK-----EEVYNHV 116
            +      S + NN++K K+E      Y+K  E+ K  +  ++++K S      E + N +
Sbjct: 193 EMIDIRISEVENNLKKLKKEKEDAEKYIKLNEELKAAKYALILKKVSYLNVLLENIQNDI 252

Query: 117 KNGDKLIEKMAFANELILKFGEVSRENQMLGLKVNSLEEKIVDLSNQPKNDEISKLRKSI 176
            KN ++L       NE + K  E+  E + L L++N+     I++  N+  N+E+ +L KSI
Sbjct: 253 KNLEEL------KNEFLSKVREIDVEIENLKLRLNN----IINELNEKGNEEVLELHKSI 302

Query: 177 SSFERELSRFEDVGYSEAEEIKSTLRRILN 206
               E E+   + V  S   E+K     I N
Sbjct: 303 KELEVEIENDKKVLDSSINELKKVEVEIEN 332
```

No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 1766 (GBS315) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 42 (lane 4; MW 26.7kDa) and in Figure 239 (lane 5; MW 41kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 47 (lane 5; MW 52kDa).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 558

A DNA sequence (GBSx0597) was identified in *S.agalactiae* <SEQ ID 1767> which encodes the amino acid sequence <SEQ ID 1768>. This protein is predicted to be surface protein. Analysis of this protein sequence reveals the following:

```
Possible site: 26
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -7.70   Transmembrane  229 - 245 ( 226 - 248)

----- Final Results -----
              bacterial membrane --- Certainty=0.4079(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAA47097 GB:X66468 orf iota [Streptococcus pyogenes]
 Identities = 90/262 (34%), Positives = 138/262 (52%), Gaps = 26/262 (9%)

Query: 4   VKVLSLITV-SGLFLMAGNLSASADVVISGGDTIMLSGVDAGVSDSIMPPPSSINPV--- 59
           +K L+L+T+ S   L++  + + AD   S  D  +L+  D  V      P + ++PV
Sbjct: 1   MKKLALLTLFSTTLLVSAPIVSFADETASSSDINILADDDPVVPVEPTDPTTPVDPVDPV 60

Query: 60  -----------TDTTEPSAPTPSTDPI--TDTTEPSAPTPSTDPI--TDTTEPSAPTPST 104
                      T+ TEP+ PT  T+P   T+ TEP+ PT  T+P   T+ TEP+ PT  T
Sbjct: 61  DPVDPVDPVDPTEPTEPTEPTEPTEPTEPTEPTEPTEPTEPTEPTEPTEPTEPTEPTEPT 120

Query: 105 DQTTGTTDSS-TPSSSTTNPVDGITDNGTKPNAGIDKPSTNKPSDHSESSI--KPVTKPT 161
           + T  T  +  T  S  T P +      T+P    +     +PS  +E ++  KPV
Sbjct: 121 EPTEPTEPTEPTEPSKPTEPTE--PSKPTEPTEPTEPSKPTEPSKPTEPTVPNKPVDTNP 178

Query: 162 INQPITTVTGDQVIGTQDGKVLVQTPSGTQLK-DAAEVGGNVQKDGTVAIKKSDGKIEVL 220
           I  P+ T TG  ++  +D K ++Q   GT  K +A E+G +VQKDGTV +K SDGK++VL
Sbjct: 179 IENPVNTDTGVVIVAVEDSKPIIQLADGTTKKVEAKEIGADVQKDGTVTVKGSDGKMKVL 238

Query: 221 PKTGEGKTI-FTIVGLLLIAGA 241
           PKTGE   I  +++G L++ G+
Sbjct: 239 PKTGETANIALSVLGSLMVLGS 260
```

There is also homology to SEQ ID 760.

SEQ ID 1768 (GBS141) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 19 (lane 4; MW 35kDa). The GBS141-His fusion product was purified (Figure 194, lane 3) and used to immunise mice. The resulting antiserum was used for FACS (Figure 295), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 559

A DNA sequence (GBSx0598) was identified in *S.agalactiae* <SEQ ID 1769> which encodes the amino acid sequence <SEQ ID 1770>. Analysis of this protein sequence reveals the following:

```
Possible site: 18
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
               bacterial outside --- Certainty=0.3000(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

A related GBS gene <SEQ ID 8601> and protein <SEQ ID 8602> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 4
McG: Discrim Score:     14.39
GvH: Signal Score (-7.5): -1.23
     Possible site: 18
>>> Seems to have a cleavable N-term signal seq.
ALOM program   count: 0 value:   8.96 threshold:  0.0
   PERIPHERAL  Likelihood =  8.96     104
 modified ALOM score:  -2.29

*** Reasoning Step: 3

----- Final Results -----
                bacterial outside --- Certainty=0.3000(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

SEQ ID 1770 (GBS17) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 4 (lane 2; MW 24kDa).

The His-fusion protein was purified as shown in Figure 189, lane 10.

Example 560

A DNA sequence (GBSx0599) was identified in *S.agalactiae* <SEQ ID 1771> which encodes the amino acid sequence <SEQ ID 1772>. Analysis of this protein sequence reveals the following:

```
Possible site: 23
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
                bacterial outside --- Certainty=0.3000(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS gene <SEQ ID 10779> and protein <SEQ ID 10780> were also identified. A further related GBS nucleic acid sequence <SEQ ID 10957> which encodes amino acid sequence <SEQ ID 10958> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 1772 (GBS643) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 129 (lane 2-4; MW 79kDa) and in Figure 186 (lane 2; MW 79kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 129 (lane 5-7; MW 54kDa) and in Figure 176 (lane 5; MW 54kDa).

GBS643-GST was purified as shown in Figure 236, lane 7.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 561

A DNA sequence (GBSx0600) was identified in *S.agalactiae* <SEQ ID 1773> which encodes the amino acid sequence <SEQ ID 1774>. Analysis of this protein sequence reveals the following:

```
Possible site: 27
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.5815(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 562

A DNA sequence (GBSx0601) was identified in *S.agalactiae* <SEQ ID 1775> which encodes the amino acid sequence <SEQ ID 1776>. This protein is predicted to be membrane protein. Analysis of this protein sequence reveals the following:

```
Possible site: 33
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-13.32   Transmembrane  311 - 327 ( 282 - 332)
   INTEGRAL    Likelihood =-10.46   Transmembrane  293 - 309 ( 282 - 310)
   INTEGRAL    Likelihood = -8.55   Transmembrane  390 - 406 ( 388 - 410)
   INTEGRAL    Likelihood = -7.64   Transmembrane   49 -  65 (  40 -  69)
   INTEGRAL    Likelihood = -5.68   Transmembrane  100 - 116 (  98 - 122)
   INTEGRAL    Likelihood = -4.35   Transmembrane  130 - 146 ( 127 - 148)
   INTEGRAL    Likelihood = -3.88   Transmembrane  344 - 360 ( 342 - 363)

----- Final Results -----
               bacterial membrane --- Certainty=0.6328(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB70618 GB:AJ243106 membrane protein [Streptococcus thermophilus]
 Identities = 234/665 (35%), Positives = 379/665 (56%), Gaps = 59/665 (8%)

Query: 13  FAKVKDVDIFALKAYMEITH-GAETGAQSILLDVFVNFPFFLLNLIVGLFSVILRFFENF 71
            FAK+K VDIF+LK+YME T+ G+  GA  ++ ++FVN  FF+LN +VG FS+++R  E
Sbjct: 5   FAKLKGVDIFSLKSYMEPTNFGSFNGAWVLINELFVNLFFFILNAVVGFFSLLIRILEKI 64

Query: 72  SLYDTYKQTVYHSSQKLWENLSGN--GSYTS-SLLYLLVAISAFSIFISYLFSKGDFSKR 128
             LY TYK  V+H +  +W   +G+  G+ T+ SL+  L+ + AF +F  Y FSKG FS+
Sbjct: 65  DLYATYKTYVFHGASSIWHGFTGSNTGNITNKSLVGTLLLVLAFYLFYQYFFSKGSFSRT 124

Query: 129 LIHLFVVIILGMGYFGTIQSTSGGIYILDTVHQLAGSFSDAVTNLSLDNPSGGKTKITQK 188
            L+H+ +V++L +GYFGT+  TSGG+Y+LDTV+ ++   +  +  + +D       KI +
Sbjct: 125 LLHVCLVLLLALGYFGTVAGTSGGLYLLDTVNNVSKDVTKKIAGIKVDYAKDKSIKIGK- 183

Query: 189 SSVADNYVMKTSYTAYLFVNTGQLNGKFHNNQTGKEEKFDNEQVLGKYDKSGKFITPKQK 248
             S++D+Y+ +TSY AY+FVNTGQ NGK+ N+Q GKEE FD+ +VLG  DK+G F   K K
```

```
Sbjct: 184 -SMSDSYIAETSYKAYVFVNTGQENGKYKNSQDGKEEAFDDSKVLGTSDKNGNFKAVKAK 242

Query: 249 DILNYTDNLGDKATEGEEKNRWLSAVNDYLWIKSGYVILKIFEAVILAVPLILIQLIAFM 308
           +   Y D+LG+ A +  EKNRW+SA+ D+++ +  YVI KI EA +LAVP+ILIQL+  +
Sbjct: 243 ERSKYLDDLGEGANDDGEKNRWVSAMPDFIFTRVFYVIFKIVEAFVLAVPIILIQLLNVV 302

Query: 309 ADVLVIILMFIFPLALLVSFLPRMQDIIFNVLKVMFGAVSFPALAGFLTLIVFYTQTLIA 368
           A +LV+ ++ +FP+ LL+SF+PRMQ+++F VLKVMFG + FPA+   LTL++FY + +I
Sbjct: 303 AQILVLTMILLFPVVLLMSFVPRMQELVFGVLKVMFGGLIFPAITTLLTLLIFYIEKMIE 362

Query: 369 TFVKKKFTDGSLLSGSNFKGQAILFMLLITVFVQGCVFWGIWKYKETFLRLIIGSRASQV 428
             V   F DG L +  +     ++F LL++V  +G +++ IW++K   L+ I+GS+A  V
Sbjct: 363 NIVTNGF-DGVLKTLPSLLLFGLVFKLLVSVVSKGVIYFLIWRFKGQLLQFILGSKARMV 421

Query: 429 -------INQSVDKINEKAENLGITPKSIYERAHDMSSLAMMGAGYGVGTMMNAQ---DN 478
                  +   V K  E A  +   P      A  + +  + GAG+G G MMNA+    N
Sbjct: 422 ATDIGTKVEHGVTKSKEVASQV---PTRSLATAQHLGNFTLAGAGFGTGVMMNAKSHFQN 478

Query: 479 WNAFKERQQANLDDGQSKTNDADKYDEANADDTVISKEAELTNEGEYQSELPKEASKRIE 538
             +F  R++ +  +    +   +      + + +I             ++ P +  K I
Sbjct: 479 AGSFFTRKEPSQPETVMPSGPTEAPITPESPEPIIP-----------PTQTPPDNFKTIG 527

Query: 539 QLGKESSYELSFISEGNSTEEILKNVKSDNHTFQEGDGDTSLTNQDMITNDIENHSNNYT 598
           +       +   +SEG  + E                            ++ +     +
Sbjct: 528 EEKPTPPSDSPIMSEGTPSSE----------------------------DEFQTLKEEWM 559

Query: 599 SPLKQRKLNKLEGELSQFNSDVSMTKNHGKNAFEKGFNASKTKEVRKQHNLERQSKVLEE 658
           SP KQ ++N LE  L  +    +M K  G NAF + +  + T++ + + N+ER+ ++ +
Sbjct: 560 SPFKQHRINTLERRLDAYKDPQAMYKAQGSNAFTRAYRKTLTRDDKIRANIERRDRLTQR 619

Query: 659 LEKLR 663
           L +LR
Sbjct: 620 LNQLR 624
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 563

A DNA sequence (GBSx0602) was identified in *S.agalactiae* <SEQ ID 1777> which encodes the amino acid sequence <SEQ ID 1778>. This protein is predicted to be conjugative protein. Analysis of this protein sequence reveals the following:

```
Possible site: 28
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3714(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB70617 GB:AJ243106 conjugative protein [Streptococcus thermophilus]
 Identities = 515/757 (68%), Positives = 612/757 (80%), Gaps = 1/757 (0%)

Query: 1   MSDFEADLADDVKELGLETLDFTVDTLTHEMEIPYQFDWLIGVDLGKGQYNANIKEFIYN 60
           M DF   LADD +ELG E L +TVD LT EMEIPYQFDW+IGV L K  + A +K+  Y
Sbjct: 78  MRDFSEALADDSRELGEELLLYTVDRLTDEMEIPYQFDWVIGVTLRKQNHGATVKDLAYE 137

Query: 61  QFESIASNFASLAGYEVEVDEDWYKEHSEEELLVYSLLSTLKAKRLTDVDLFYYQRMQFL 120
            F   +   A   GYE  +   WY ++  +E  ++   S L+AKRLT+ +LFYYQRMQ+L
Sbjct: 138 SFNEFSEKIAKGLGYEYALSPTWYDDYRSDEFTIFQAFSVLRAKRLTNEELFYYQRMQYL 197
```

```
Query: 121 RYVPHTKSEVIANRNMLNVTDTLIKSLEGGFLKLESAYGSSFVSVLPVGRFSTIFNGFHL 180
           RY+PH K EV+ANR+  N+TDTLIK L+GGFL+LES YGSSFV++LPVG+F   FNGFHL
Sbjct: 198 RYIPHYKKEVLANRSQFNITDTLIKVLKGGFLELESPYGSSFVTILPVGKFPVQFNGFHL 257

Query: 181 GELVQRMSFPVELRFKAEFIDKTKLGGTMGRSNTRYDQIMKEAYNTNTVQQDDILMGAYS 240
           GE VQR++FPVELR KAEFID  K+ G MGRSNTRY  IM+EA NT+TVQQD+I+MG+ S
Sbjct: 258 GEFVQRLNFPVELRIKAEFIDTNKIKGRMGRSNTRYRNIMEEAENTDTVQQDEIIMGSIS 317

Query: 241 LKDLMKKVGNKEEIIEYGCYLVVAGSSLNQLKQRRYAILSYFDDMKVNVYEASHDTPYLF 300
           LKDLMKKVGNKE+IIEYG YL+V+ SS+NQL+QRR  IL+YFDDM V + EAS D PYLF
Sbjct: 318 LKDLMKKVGNKEDIIEYGAYLIVSASSVNQLRQRRQVILNYFDDMGVEISEASQDGPYLF 377

Query: 301 QALLYGQDLQKTTRKWNHLVTARGFSELMLFTNTQSGNRIGWYIGRVDNRLTAWDSIDEA 360
           QALLYG++LQK TR W H+VTARGFSELM FTNT SGNRIGWYIGRVDN +  WDSI +A
Sbjct: 378 QALLYGENLQKKTRTWTHMVTARGFSELMPFTNTSSGNRIGWYIGRVDNWIGRWDSIAKA 437

Query: 361 IMGSKNLVLFNATVANKEDVAGKVTKNPHVIITGATGQGKSYLAQMIFLHTAQQNVRVLY 420
           I  SKN+VL+NATV NKED+AGK+TKNPH+IITGATGQGKS+LAQ+IFL  A QNV+ LY
Sbjct: 438 IDSSKNIVLYNATVGNKEDIAGKITKNPHIIITGATGQGKSFLAQIIFLSVALQNVKTLY 497

Query: 421 VDPKRELRQHYLKVVSDPEYARKFPLRKKQIEETNFVTLDSSVKENHGVLDPIVILDKEG 480
           +DPKRELR HY +V++ PE+AR++P RKKQI+  NFVTLDSS+  NHGVLDPIV+LDKE
Sbjct: 498 IDPKRELRNHYQEVINSPEFARRYPERKKQIDNFNFVTLDSSLPSNHGVLDPIVVLDKEQ 557

Query: 481 ASSTAKNMLLYLLKNATEIKLDQTTALTEAISQVIAKREAGEVVGFNQVIEVLIDSESDE 540
           A   AKNML +LL+   ++ +DQ TA+TEAI+ ++ +R AGE VGF  V+E L ++ S E
Sbjct: 558 AVEVAKNMLEFLLQAVDDVTMDQKTAITEAINTIVERRVAGENVGFKHVLETLRNASSSE 617

Query: 541 VQSVGRYFKAIIQNSILELAFSDGDVAGLSYEERVTVLEVADLSLPKDGSDHISDHESNS 600
           + SVGRY  +I+ NSILELAFSDG   GL+YE RVT+LEV +L LPKD S  ISDHE NS
Sbjct: 618 IASVGRYLTSIVTNSILELAFSDGTTPGLNYESRVTILEVNNLKLPKDDSTKISDHERNS 677

Query: 601 IALMFALGAFCKHFGERSDDE-TVEIFDEAWVLMQSSEGKAVIKSMRRVGRSKYNVLMLV 659
           IALMFALGAFC HFGER+++E T+E FDEAW+LM+S+EGKAVIK+MRR+GRSK N L L+
Sbjct: 678 IALMFALGAFCTHFGERNENEDTIEFFDEAWILMKSAEGKAVIKNMRRIGRSKNNTLALI 737

Query: 660 SQSVHDAENDDDTTGFGTIFSFYEKSEREDILSHVGLEVTPKNLEWIDNMISGQCLYYDV 719
           +QSVHDAENDDDTTGFGTIF+FYEKSEREDIL HV LEVT  NLEWIDNMISGQCLYYDV
Sbjct: 738 TQSVHDAENDDDTTGFGTIFAFYEKSEREDILRHVNLEVTESNLEWIDNMISGQCLYYDV 797

Query: 720 YGNLNMISIHNIHPDIDPLLKPMKKTVSSHLENKYAS 756
           YGNLNMIS+HN+  DID LLKPMK TVSS LENKYAS
Sbjct: 798 YGNLNMISVHNLFEDIDMLLKPMKATVSSSLENKYAS 834
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 564

A DNA sequence (GBSx0604) was identified in *S.agalactiae* <SEQ ID 1779> which encodes the amino acid sequence <SEQ ID 1780>. This protein is predicted to be ISL2 protein. Analysis of this protein sequence reveals the following:

```
Possible site: 26
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3469(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAC18595 GB:AJ278419 IS1381 transposase [Streptococcus pneumoniae]
 Identities = 110/125 (88%), Positives = 119/125 (95%)

Query: 81  MNYEASKQLTDVRFKRLVGVQRTTFEEMLAVLKTAYQRKHAKGGRTPKLSLEDLLMATLQ 140
            MNYEASKQLTD RFKRLVGVQRTTFEEMLAVLKTAYQ KHAKGGR PKLSLEDLLMATLQ
Sbjct: 1   MNYEASKQLTDARFKRLVGVQRTTFEEMLAVLKTAYQLKHAKGGRKPKLSLEDLLMATLQ 60

Query: 141 YMREYRTYEQIAADFGIHESNLIRRSQWVESTLIQSGFTISKTHLSAEDTVIVDATEVKI 200
            Y+REYRTYE+IAADFG+HESNL+RRSQWVE TL+QSG TIS+T LS+EDTV++DATEVKI
Sbjct: 61  YVREYRTYEEIAADFGVHESNLLRRSQWVEVTLVQSGVTISRTPLSSEDTVMIDATEVKI 120

Query: 201 NRPKK 205
            NRPKK
Sbjct: 121 NRPKK 125
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 565

A DNA sequence (GBSx0605) was identified in *S.agalactiae* <SEQ ID 1781> which encodes the amino acid sequence <SEQ ID 1782>. Analysis of this protein sequence reveals the following:

```
Possible site: 61
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-12.58    Transmembrane    39 -   55 (   32 -   66)

----- Final Results -----
               bacterial membrane --- Certainty=0.6031(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 566

A DNA sequence (GBSx0606) was identified in *S.agalactiae* <SEQ ID 1783> which encodes the amino acid sequence <SEQ ID 1784>. This protein is predicted to be Cag-W. Analysis of this protein sequence reveals the following:

```
Possible site: 59
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -3.82    Transmembrane    50 -   66 (   49 -   66)
   INTEGRAL    Likelihood = -3.72    Transmembrane    25 -   41 (   23 -   45)

----- Final Results -----
               bacterial membrane --- Certainty=0.2529(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 567

A DNA sequence (GBSx0607) was identified in *S.agalactiae* <SEQ ID 1785> which encodes the amino acid sequence <SEQ ID 1786>. Analysis of this protein sequence reveals the following:

```
Possible site: 55
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -7.80   Transmembrane   36 -  52 (  32 -  60)

----- Final Results -----
               bacterial membrane --- Certainty=0.4121(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB12298 GB:Z99106 similar to transposon protein [Bacillus subtilis]
 Identities = 68/339 (20%), Positives = 133/339 (39%), Gaps = 49/339 (14%)

Query: 16  KKEEGGKQPKTKEVKQRTANFIV--YGILGLLFIVGFFGSLRAIGLSNQVQHLKETVIAV 73
           K+ E  ++ K K  + R+    V  +  +G L  +     L +I   +Q+  +K+
Sbjct: 24  KRIERPEKDKQKVPRDRSKLIAVTLWSCVGSLLFICLLAVLLSINTRSQLNDMKDETNKP 83

Query: 74  EKKSKHKKTDDSLDISRIQYYMNNFVYYYINYS--QDTADQRKTELENY--------YSF 123
               K K     + ++  + +++ F+  Y+N    Q++ ++R   LE+Y        +
Sbjct: 84  TNDDKQK-----ISVTAAENFLSGFINEYMNVKNDQESIEKRMQSLESYMVKQEDNHFED 138

Query: 124 STASMTDDVRKSRTLQTQRLISVEKEKDYYIALMRIGYEV-------------------- 163
                 D ++  R L+   L +V++     +   ++ YE
Sbjct: 139 EERFNVDGLKGDRELKGYSLYNVKEGDKNSLFQYKVTYENLYPVEKEVEKEVKDGKKKKK 198

Query: 164 --------DKKSYQMNLAVPFQMQRGLLAIVSQPYTVAEDLYLGKSKAFEKKTLDQVKEL 215
                   +K   QM L +P   +    A+ + PY     +Y  K     K   +   E
Sbjct: 199 VKEKVKTNEKYEKQMLLNIPVTNKGDSFAVSAVPYFT--QIYDLKGDIAFKGKEETRDEY 256

Query: 216 SKEQVSSIQKFLPVFFNKYALINKTDLKLLMKTPELMGKGFKVSELDLNNAIYYQEKKHQ 275
           + E+  SI+ FL  FF KYA   K ++  +MK PE +       E  + +   ++ KK
Sbjct: 257 AGEKKESIESFLQNFFEKYASEKKEEMVYMMKKPEALEGNLLFGE--VQSVKIFETKKGF 314

Query: 276 VVQLSVTFEDLVTGGTRSENFTLYLFKADNGWYVEEMYH 314
            V  +V F++       +E F+L + +    +YV ++ H
Sbjct: 315 EVFCAVRFKEKENDIPVNEKFSLEITENSGQFYVNKLKH 353
```

No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 1786 (GBS333d) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 145 (lane 8-10; MW 58kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 145 (lane 11 & 13; MW 33kDa), in Figure 182 (lane 2; MW 33kDa) and in Figure 185 (lane 3; MW 58kDa).

GBS333d-GST was purified as shown in Figure 236, lane 2.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 568

A DNA sequence (GBSx0608) was identified in *S.agalactiae* <SEQ ID 1787> which encodes the amino acid sequence <SEQ ID 1788>. Analysis of this protein sequence reveals the following:

```
Possible site: 54
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4177(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB38326 GB:Y17736 hypothetical protein [Streptomyces
          coelicolor A3(2)]
 Identities = 45/80 (56%), Positives = 56/80 (69%)

Query: 4  FTEEAWKDYVSWQQEDKKILKRINRLIEDIKRDPFEGIGKPEPLKYHYSGAWSRRITEEH 63
          FT   W+DYV W + D+K+ KRINRLI DI RDPF+G+GKPEPLK   SG WSRRI + H
Sbjct: 5  FTSHGWEDYVHWAESDRKVTKRINRLIADIARDPFKGVGKPEPLKGDLSGYWSRRIDDTH 64

Query: 64 RLIYMIEDGEIYFLSFRDHY 83
          RL+Y   D ++  +  R HY
Sbjct: 65 RLVYKPTDDQLVIVQARYHY 84
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 569

A DNA sequence (GBSx0609) was identified in *S.agalactiae* <SEQ ID 1789> which encodes the amino acid sequence <SEQ ID 1790>. Analysis of this protein sequence reveals the following:

```
Possible site: 53
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.5669(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10211> which encodes amino acid sequence <SEQ ID 10212> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAD17306 GB:AF121418 putative Phd protein [Francisella
          tularensis subsp. novicida]
 Identities = 26/84 (30%), Positives = 45/84 (52%)

Query: 4  MEAIVYSHFRNNLKDYMKKVNDEFEPLIVVNKNPDENIVVLSQDSWESLQETIRLMENDY 63
          M+ + YS FRN L D M +V     P+IV   +  E +V++S + +++ +ET  LM +
Sbjct: 1  MQTVNYSTFRNELSDSMDRVTKNHSPMIVTRGSKKEAVVMMSLEDFKAYEETAYLMRSMN 60

Query: 64 LSHKVINGISQVKEKQVTKHGLIE 87
             ++ N I +V+     +  LIE
Sbjct: 61 NYKRLQNSIDEVESGLAIQKELIE 84
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 570

A DNA sequence (GBSx0610) was identified in *S.agalactiae* <SEQ ID 1791> which encodes the amino acid sequence <SEQ ID 1792>. Analysis of this protein sequence reveals the following:

```
Possible site: 55
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2407(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 571

A DNA sequence (GBSx0611) was identified in *S.agalactiae* <SEQ ID 1793> which encodes the amino acid sequence <SEQ ID 1794>. Analysis of this protein sequence reveals the following:

```
Possible site: 28
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1274(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10213> which encodes amino acid sequence <SEQ ID 10214> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAB60015 GB:U09422 ORF18 [Enterococcus faecalis]
 Identities = 41/140 (29%), Positives = 73/140 (51%), Gaps = 3/140 (2%)

Query: 23  FPVEMSELKLALGLREEDDLEYIIADSDCQL-LKEHDSIEMINQFVELVENVDSELVKAV 81
            FP++  E+K  +GL +E + EY I D +    + E+ SI  +N+  E+V  +  EL   +
Sbjct: 26  FPIDFEEVKEKIGLNDEYE-EYAIHDYELPFTVDEYTSIGELNRLWEMVSELPEELQSEL 84

Query: 82  HQVIGYTASDFVDYDFNFGDCCLLSDVTTRRELGEYYFDELGVQGVGKEALEMYFDHEAY 141
              ++ + +S   +   +  D  + SD     ++  YY +E G  G    +L+ Y D++AY
Sbjct: 85  SALLTHFSS-IEELSEHQEDIIIHSDCDDMYDVARYYIEETGALGEVPASLQNYIDYQAY 143

Query: 142 GRDIDLESQGGFSDYGYVEI 161
            GRD+DL      +++G  EI
Sbjct: 144 GRDLDLSGTFISTNHGIFEI 163
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 572

A DNA sequence (GBSx0612) was identified in *S.agalactiae* <SEQ ID 1795> which encodes the amino acid sequence <SEQ ID 1796>. Analysis of this protein sequence reveals the following:

```
Possible site: 31
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1366(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 573

A DNA sequence (GBSx0613) was identified in *S.agalactiae* <SEQ ID 1797> which encodes the amino acid sequence <SEQ ID 1798>. Analysis of this protein sequence reveals the following:

```
Possible site: 41
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1484(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 574

A DNA sequence (GBSx0614) was identified in *S.agalactiae* <SEQ ID 1799> which encodes the amino acid sequence <SEQ ID 1800>. This protein is predicted to be abortive phage resistance protein. Analysis of this protein sequence reveals the following:

```
Possible site: 58
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2205(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10215> which encodes amino acid sequence <SEQ ID 10216> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAB53710 GB:U94520 abortive phage resistance protein
           [Lactococcus lactis]
```

```
 Identities = 131/499 (26%), Positives = 210/499 (41%), Gaps = 97/499 (19%)

Query: 3   MFSKIEFKNFMSFSNLT------------------FDLLNRGKCKDIIAIYGENGSGKTN 44
            M     F+NF+SF  L+                   D+ N  K   +  IYG N SGK++
Sbjct: 1   MLVNFRFENFLSFDKLSTFSMAPGKSRQHMEDLIELDIKNNQKLLKLSTIYGANASGKSS 60

Query: 45  IVEAF---KLLVL-----SLQSMESLNENTRLQSLLKEQTNKE---ENQKTNFGDISEIL 93
             V+A    K L++      L    S N+NT   SL + +   E   E++  ++G  S IL
Sbjct: 61  FVDAIGISKSLIIRGFYNGLVLSNSYNKNTVDNSLNETKFEYEIVIEDKVYSYG-FSVIL 119

Query: 94  DKISFFTTFKGIAKNTHRIASEGNTILKYYFNIEKDNGYYLLEYNENNELVKEELVFKIK 153
                F + +     N ++           Y    KDN      YN N+E     L    +
Sbjct: 120 SLKKFMSEWLYDITNDEKM---------IYTIDRKDN-----SYNINDEF----LNLDEQ 161

Query: 154 SNKGVHFSITNIDGLSQSLNKTIFKNTIFKDLTEQIEKYWGKHTFLSIFN--NYCLEV-- 209
            SN  +   I +    S + N  +F N++  D  + IE       F  +FN  N  LEV
Sbjct: 162 SNNRISIYIDD----SANDNTQLFLNSL-NDGKKTIESKDNSTIFKKVFNWFNNTLEVLG 216

Query: 210 ---------------NEEF---INEQVSINFQKVVDEFDKIFIWSGNFRGPFHSTELLLK 251
                            EEF   + + + +N   V+D          N   P    E +L
Sbjct: 217 PGDEARGSIASLTQEEEEFKEDLGKYLELNDTGVIDIVQVPVDNLSNV--PAKLQERILD 274

Query: 252 DISKGKIDKSEKEKLSYTEEIIYKYFSALYIDIKDVKYKQDAQGQEIKYELMIRKNIGGD 311
            +I+   I K +KE+     E I   F+ +    +++   Q+   Q   +EL   K+  G
Sbjct: 275 NITT-DIKKKKKER-----EDIEISFNTILNTSQNIYIIQNNDEQFEYFELKF-KHKNGT 327

Query: 312 LLDVPISLESQGTKNLLDLLKV-FNNVLDGKICIVDEIDSGIHDLLMNSILNDLK--GSV 368
            L    +S ES GT  L++L  V F+N  D K+ ++DEID  +H LL  + +   K   S+
Sbjct: 328 LYS--LSEESDGTVRLIELFSVLFHN--DEKVFVIDEIDRSLHPLLTYNFIESFKKQKSI 383

Query: 369 NGQLIFTTHDTTLL--KELSPSSAYFLNVDIKGNKVIISGNEADKKIGVNNNLEKLYLSG 426
            N QLI TTH+  +L  + L     +F++ + +GN  + S  E  ++   + ++   YL+G
Sbjct: 384 N-QLIVTTHEDYILNFELLRRDEVWFVDKNFEGNSSMFSLEEFKERF--DKDINTSYLNG 440

Query: 427 FFGAVPDPLDIDFSDLFLD 445
             +G +P+ L   FS+   D
Sbjct: 441 RYGGIPN-LSCLFSEFAKD 458
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 575

A DNA sequence (GBSx0615) was identified in *S.agalactiae* <SEQ ID 1801> which encodes the amino acid sequence <SEQ ID 1802>. This protein is predicted to be repressor (rstR-1). Analysis of this protein sequence reveals the following:

```
Possible site: 37
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3724(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAB84427 GB:AF027868 transcription regulator [Bacillus subtilis]
 Identities = 31/81 (38%), Positives = 53/81 (65%), Gaps = 2/81 (2%)

Query: 9   QKLKELRKEKKLTQTELASKLNISQKSYSNWESGKAEPTLDNIIKLANILDVTVDYLLGR 68
            Q+L++LRK  KLT  +LA K+ I++ SY  +E+   +P LD ++ LA + DV+VDY+LG
Sbjct: 4   QRLRQLRKAHKLTMEQLAEKIGIAKSSYGGYEAESKKPPLDKLVILARLYDVSVDYILGL 63
```

```
Query: 69 SDNFSNTIVLSKNNMKSFSKR 89
          +D+    +   + N+K F ++
Sbjct: 64 TDDPDPKV--ERKNLKEFLEK 82
```

There is also homology to SEQ ID 1740.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 576

A DNA sequence (GBSx0616) was identified in *S.agalactiae* <SEQ ID 1803> which encodes the amino acid sequence <SEQ ID 1804>. Analysis of this protein sequence reveals the following:

```
Possible site: 13
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.3607(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 577

A DNA sequence (GBSx0617) was identified in *S.agalactiae* <SEQ ID 1805> which encodes the amino acid sequence <SEQ ID 1806>. Analysis of this protein sequence reveals the following:

```
Possible site: 22
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.0564(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10217> which encodes amino acid sequence <SEQ ID 10218> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB12294 GB:Z99106 similar to transposon protein [Bacillus subtilis]
 Identities = 93/348 (26%), Positives = 164/348 (46%), Gaps = 28/348 (8%)

Query: 81  SRLQVMIDYVRITLKDVRDLEFFCRNFLHCAFKEFQPFESKLMNYNHLWKRGDIWIFDFA 140
            S L  M+DY+R++ K   D++      LH +       +S    Y   ++   I +F  A
Sbjct: 26  SPLVSMVDYIRVSFK-THDVDRIIEEVLHLSKDFMTEKQSGFYGYVGTYELDYIKVFYSA 84

Query: 141 DKHETGNFQITVQLSGRGCRQLELLMETEKFTWHDWLSYLRNSYRDDMNVTRFDIAIDEL 200
                 G   + +++SG+GCRQ E  +E  K TW+D   + ++  +   + TRFD+AID+
Sbjct: 85  PDDNRG---VLIEMSGQGCRQFESFLECRKKTWYD---FFQDCMQQGGSFTRFDLAIDD- 137

Query: 201 YLGKDRENEQFHLSDMISKYYRHELDFESLRTWNYIGGGSLNFSDMEEIEQNRQGISLYF 260
                  +   F + +++ K  + E      R  ++   GS + SD         G ++YF
Sbjct: 138 ------KKTYFSIPELLKKAQKGEC-ISRFRKSDF--NGSFDLSD-----GITGGTTIYF 183
```

```
Query: 261 GSRQSEMYFNFYEKRYEIAKQEGITVEEALEIFELWNRYEIRLSQSKANAAVDEFISGVP 320
           GS++SE Y  FYEK YE A++  I +EE  +    WNRYE+RL   +A  A+D  +
Sbjct: 184 GSKKSEAYLCFYEKNYEQAEKYNIPLEELGD----WNRYELRLKNERAQVAIDALLKTKD 239

Query: 321 IGEISRGLIVSKIDVYDGKNEY--GSFQADRKWQLMFGGVEPLKFVTKPEAYSIERTLRW 378
           +  I+  +I + +   D         ++    W    G V  L    KP+    +++  W
Sbjct: 240 LTLIAMQIINNYVRFVDADENITREHWKTSLFWSDFIGDVGRLPLYVKPQKDFYQKSRNW 299

Query: 379 LSDSVSPSLAMIREYDMIVDGDYLQTILNSGEVNERGEKILDSIKASL 426
           L +S +P++ M+ E D  +    L  ++   E+ ++ +K+LD   A +
Sbjct: 300 LRNSCAPTMKMVLEADEHLGKTDLSDMIAEAELADKHKKMLDVYMADV 347
```

No corresponding DNA sequence was identified in *S.pyogenes*.

A related GBS gene <SEQ ID 8603> and protein <SEQ ID 8604> were also identified. Analysis of this protein sequence reveals a RGD motif at residues 131-133.

The protein has homology with the following sequences in the databases:

```
29.4/54.5% over 342aa
                                                          Bacillus subtilis
  EGAD|108511| hypothetical protein Insert characterized OMNI|NT01BS0566 conserved
hypothetical protein Insert characterized
  GP|1881297|dbj|BAA19324.1||AB001488 SIMILAR TO ORF20 OF ENTEROCOCCUS FAECALIS TRANSPOSON
TN916. Insert characterized
  GP|2632787|emb|CAB12294.1||Z99106 similar to transposon protein Insert characterized
  PIR|G69774|G69774 transposon-related protein homolog ydcR - Insert characterized

ORF00101(205 - 1581 of 1887)
EGAD|108511|BS0487(6 - 348 of 352) hypothetical protein {Bacillus subtilis} OMNI|NT01BS0566
conserved hypothetical protein GP|1881297|dbj|BAA19324.1||AB001488 SIMILAR TO ORF20 OF
ENTEROCOCCUS FAECALIS TRANSPOSON TN916. {Bacillus subtilis}
GP|2632787|emb|CAB12294.1||Z99106 similar to transposon protein {Bacillus subtilis}
PIR|G69774|G69774 transposon-related protein homolog ydcR - Bacillus subtilis
%Match = 9.7
%Identity = 29.3  %Similarity = 54.4
Matches = 103  Mismatches = 146  Conservative Sub.s = 88

153       183       213       243       273                 489       519
GVV*RSENGHAGHSAHTRALQRDLSILKPPFSNRGVRNEKFRILTPKNLYVSRVFR~~~~EQGKRKLTLEKYQEIKSHFG
                         : :|| :||||
                      MDELKQPPHANRGV--------------------~~~---------------------
                              10

567       597       627       657       687       717       747
YLV--ENDS--SRLQVMIDYVRITLKDVRDLEFFCRNFLHCAFKEFQPFESKLMNYNHLWKRGDIWIFDFADKHETGNFQ
 :|  :|::  | |  |:||:|:::|   |::      || :       :| :  |   ::   | :|  |     |
VIVKEKNEAVESPLVSMVDYIRVSFK-THDVDRIIEEVLHLSKDFMTEKQSGFYGYVGTYELDYIKVFYSAPDDNRG---
              30        40         50        60        70        80        90

777       807       837       867       897       927       957       987
ITVQLSGRGCRQLELLMETEKFTWHDWLSYLRNSYRDDMNVTRFDIAIDELYLGKDRENEQFHLSDMISKYYRHELDFES
: :::||:||||:| ::|  | ||:|   ::::  :   : ||||:|||      |:    | : ::: |  : |
VLIEMSGQGCRQFESFLECRKKTWYD---FFQDCMQQGGSFTRFDLAID------DK-KTYFSIPELLKKAQKGEC-ISR
       100       110          120       130              140       150

1017      1047      1077      1107      1137      1167      1197      1227
LRTWNYIGGGSLNFSDMEEIEQNRQGISLYFGSRQSEMYFNFYEKRYEIAKQEGITVEEALEIFELWNRYEIRLSQSKAN
:|  ::   ||:::||         | ::||||::|| |: |||| || |::  | :||    :  |||||:||   :|
FRKSDF--NGSFDLSD-----GITGGTTIYFGSKKSEAYLCFYEKNYEQAEKYNIPLEE----LGDWNRYELRLKNERAQ
          170            180       190       200       210           220

1257      1287                1341      1371      1401      1431      1461
AAVDEFISGVPIGEISRGLIVSKIDVYDG-KN-EYGSFQADRKWQLMFGGVEPLKFVTKPEAYSIERTLRWLSDSVSPSL
 |:| ::    :  |:  :| : :   |  :|     ::    |    | |  | :  ||:    :::  || :| :|::
VAIDALLKTKDLTLIAMQIINNYVRFVDADENITREHWKTSLFWSDFIGDVGRLPLYVKPQKDFYQKSRNWLRNSCAPTM
```

```
         240       250       260       270       280       290       300

1491      1521      1551      1581      1611      1641      1671      1701
AMIREYDMIVDGDYLQTILNSGEVNERGEKILDSIKASLGIL*EVSFVLYSNREFAYCVNRRNNLDKMIDLLVFMIPDRE
 |: | |  :    |  ::   |: :: :|:||   | :  :
KMVLEADEHLGKTDLSDMIAEAELADKHKKMLDVYMADVADMVV
         320       330       340       350
```

SEQ ID 8604 (GBS294) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 167 (lane 6 & 7; MW 65kDa – thioredoxin fusion), in Figure 238 (lane 2; MW 65kDa) and in Figure 40 (lane 6; MW 37kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 47 (lane 3; MW 76kDa).

Purified Thio-GBS294-His is shown in Figure 244, lane 2.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 578

A DNA sequence (GBSx0618) was identified in *S.agalactiae* <SEQ ID 1807> which encodes the amino acid sequence <SEQ ID 1808>. Analysis of this protein sequence reveals the following:

```
Possible site: 40
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -3.61   Transmembrane   24 -  40 (  20 -  41)
   INTEGRAL    Likelihood = -1.97   Transmembrane   53 -  69 (  52 -  72)

----- Final Results -----
               bacterial membrane --- Certainty=0.2444(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAB60012 GB:U09422 ORF21 [Enterococcus faecalis]
 Identities = 136/473 (28%), Positives = 228/473 (47%), Gaps = 40/473 (8%)

Query: 9   RGIKVKPYMRYMSYYL-FSFLFILFLTPVGVYSYYYLDL-------LKMMDKMSM----I 56
           RG +++P  + + ++   + L  +FL  VG++    +         L   DK+ +    I
Sbjct: 4   RGKRIRPSGKDLVFHFTIASLLPVFLLVVGLFHVKTIQQINWQDFNLSQADKIDIPYLII 63

Query: 57  SVGTGLFLAFFVSWYLTWFLQEANPLFNKLDRLKRMSKFLYENGYVYEKR-------KKS 109
           S    + +   V++    F +       +L   ++++K + EN +   ++       K S
Sbjct: 64  SFSVAILICLLVAFV---FKRVRYDTVKQLYHRQKLAKMILENKWYESEQVKTEGFFKDS 120

Query: 110 NKKTKTKYR-FPKVYVKQGKYDLSVSFEMAGGKFQKKFKDIGGELEDTFFMDFMEKTDDP 168
             +TK K   FPK+Y +     + +  E+  GK+Q +   +  +LE   + +  +K
Sbjct: 121 AGRTKEKITYFPKMYYRLKNGLIQIRVEITLGKYQDQLLHLEKKLESGLYCELTDKELKD 180

Query: 169 RFKIYKLAYSAFLSRITVKDVIWNKDKGIKLMDGYYWDFINDPHLLVAGGTGGGKTVLLR 228
            +  Y L Y    SRI++ D +  KD  ++LM   +W++   PH+L+AGGTGGGKT  +
Sbjct: 181 SYVEYTLLYDTIASRISI-DEVEAKDGKLRLMKNVWWEYDKLPHMLIAGGTGGGKTYFIL 239

Query: 229 SILRCLAEI-GVCDICDPKRADFVTMSDLSAFEGRIAFEKADIIEKFENAVTIMFARYDF 287
           +++  L        I DPK AD   ++DL +    + + K D++   E     M  R +
Sbjct: 240 TLIEALLHTDSKLYILDPKNAD---LADLGSVMANVYYRKEDLLSCIETFYEEMMKRSE- 295

Query: 288 VRNEMKRLGHKDMKKFYDY-GLEPYFFVCDEYNALMSSLSYQEREIVDNAFTQYILLGRQ 346
              EMK++ +    K Y Y GL  +F + DEY A M  L  +E   V N   Q ++LGRQ
Sbjct: 296 ---EMKQMKNYKTGKNYAYLGLPAHFLIFDEYVAFMEMLGTKENTAVMNKLKQIVMLGRQ 352

Query: 347 VGCNAIIAMQKPSADDLPTKIRSNMMHHISVGRLDDGGYVMMFGDENRNKEFRFIKYLAG 406
            G   I+A Q+P A  L   IR      +++GR+ + GY MMFG + + K+F F+K
```

```
Sbjct: 353 AGFFLILACQRPDAKYLGDGIRDQFNFRVALGRMSEMGYGMMFGSDVQ-KDF-FLK---- 406

Query: 407 RRVYGRGYSAVFGEVAREFYSPLLPKNFSFYDAFEKINRHENPFDPTENQEVS 459
            R+ GRGY  V   V  EFY+PL+PK + F +  +K++        T   EV+
Sbjct: 407 -RIKGRGYVDVGTSVISEFYTPLVPKGYDFLEEIKKLSNSRQSTQATCEAEVA 458
```

No corresponding DNA sequence was identified in *S.pyogenes*.

A related GBS gene <SEQ ID 8605> and protein <SEQ ID 8606> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop Possible site: -1   Crend: 8
McG: Discrim Score:    -10.05
GvH: Signal Score (-7.5): -3.42
     Possible site: 40
>>> Seems to have no N-terminal signal sequence
ALOM program   count: 2 value:  -3.61 threshold:  0.0
   INTEGRAL    Likelihood = -3.61   Transmembrane   24 -  40 (  20 -  41)
   INTEGRAL    Likelihood = -1.97   Transmembrane   53 -  69 (  52 -  72)
   PERIPHERAL  Likelihood =  1.01     224
 modified ALOM score:   1.22

*** Reasoning Step: 3

----- Final Results -----
              bacterial membrane --- Certainty=0.2444(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
29.9/52.7% over 456aa
                                                                           Enterococcus faecalis
  EGAD|17035| hypothetical protein Insert characterized
  GP|532554|gb|AAB60012.1||U09422 ORF21 Insert characterized

ORF00100(319 - 1677 of 2316)
EGAD|17035|17250(2 - 458 of 461) hypothetical protein {Enterococcus faecalis}
GP|532554|gb|AAB60012.1||U09422 ORF21 {Enterococcus faecalis}
%Match = 11.2
%Identity = 29.9  %Similarity = 52.7
Matches = 135  Mismatches = 199  Conservative Sub.s = 103

207       237       267       297       327       357       384       414
FQVVCLKFLHHHLRKRMLQIMETHQKMKHLKLINKR*RRGNLARLIPQYRGIKVKPYMRYMSYYL-FSFLFILFLTPVGV
                                               : || :::|  : : :::  : |: :||  ||:
                                              MKQRGKRIRPSGKDLVFHFTIASLLPVFLLVVGL
                                                      10        20        30

426       453       483       513                 570       600
Y------SYYYLDL-LKMMDKMSMISVGTGLFLAFFVSWYLTWFLQEAN-PLFNKLDRLKRMSKFLYENGYWYEKR----
:         : |: |   ||: :  :   : :| ::   : : ::        :|   ::::| : || : ||
FHVKTIQQINWQDFNLSQADKIDIPYLIISFSVAILICLLVAFVFKRVRYDTVKQLYHRQKLAKMILENKW-YESEQVKT
              50        60        70        80        90       100        110

636       663       693       723       753       783       813       843
----KKSNKKTKTKYR-FPKVYVKQGKYDLSVSFEMAGGKFQKKFKDIGGELEDTFFMDFMEKTDDPRFKIYKLAYSAFL
    | |  :|| |   |||:| :     : :  |:  ||:| ::  :  :||  :: :: :|     :  | | |
EGFFKDSAGRTKEKITYFPKMYYRLKNGLIQIRVEITLGKYQDQLLHLEKKLESGLYCELTDKELKDSYVEYTLLYDTIA
              130       140       150       160       170       180       190

873       903       933       963       993      1020      1050      1080
SRITVKDVIWNKDKGIKLMDGYYWDFINDPHLLVAGGTGGGKTVLLRSILRCLAEI-GVCDICDPKRADFVTMSDLSAFE
|||:: | :  ||  ::||   :|::   ||:|:||||||||| :: :::  |        | ||| ||   ::|| :
SRISI-DEVEAKDGKLRLMKNVWWEYDKLPHMLIAGGTGGGKTYFILTLIEALLHTDSKLYILDPKNAD---LADLGSVM
               210       220       230       240       250       260
```

```
1110      1140      1170      1200      1227      1257      1287      1317
GRIAFEKADIIEKFENAVTIMFARYDFVRNEMKRLGHKDMKKFYDY-GLEPYFFVCDEYNALMSSLSYQEREIVDNAFTQ
  : : | |::   |     |  |      |||:: :    | | | ||  :|:: ||| |:|  |  :|   | | : |
ANVYYRKEDLLSCIETFYEEMMKR----SEEMKQMKNYKTGKNYAYLGLPAHFLIFDEYVAFMEMLGTKENTAVMNKLKQ
        280       290           300       310       320       330       340

1347      1377      1407      1437      1467      1497      1527      1557
YILLGRQVGCNAIIAMQKPSADDLPTKIRSNMMHHISVGRLDDGGYVMMFGDENRNKEFRFIKYLAGRRVYGRGYSAVFG
 ::|||| |   |:| |:| |  |   ||      :::||: : || ||||| :   |:| |:| : ||     ||  |
IVMLGRQAGFFLILACQRPDAKYLGDGIRDQFNFRVALGRMSEMGYGMMFGSD-VQKDF-FLKRIKGR-----GYVDVGT
            360       370       380       390        400        410

1587      1617      1647      1677      1707      1737      1767      1797
EVAREFYSPLLPKNFSFYDAFEKINRHENPFDPTENQEVSKAILKDESLREFVEKTSENELLKGSVGFDFDDEMEEMENM
 |  |||:||:|| : | :  :|::        |   ||:
SVISEFYTPLVPKGYDFLEEIKKLSNSRQSTQATCEAEVAGVD
         430       440       450       460
```

SEQ ID 8606 (GBS216) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 42 (lane 3; MW 66.6kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 47 (lane 2; MW 91kDa).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 579

A DNA sequence (GBSx0619) was identified in *S.agalactiae* <SEQ ID 1809> which encodes the amino acid sequence <SEQ ID 1810>. Analysis of this protein sequence reveals the following:

```
Possible site: 14
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4095(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 580

A DNA sequence (GBSx0620) was identified in *S.agalactiae* <SEQ ID 1811> which encodes the amino acid sequence <SEQ ID 1812>. Analysis of this protein sequence reveals the following:

```
Possible site: 28
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0944(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10219> which encodes amino acid sequence <SEQ ID 10220> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 581

A DNA sequence (GBSx0621) was identified in *S.agalactiae* <SEQ ID 1813> which encodes the amino acid sequence <SEQ ID 1814>. Analysis of this protein sequence reveals the following:

```
Possible site: 60
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -4.94   Transmembrane  810 - 826 ( 808 - 830)

----- Final Results -----
                  bacterial membrane --- Certainty=0.2975(Affirmative) < succ>
                   bacterial outside --- Certainty=0.0000(Not Clear) < succ>
                 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
!GB:D90354 surface protein antigen precursor [Strept...

>GP:BAA14368 GB:D90354 surface protein antigen precursor
             [Streptococcus sobrinus]
 Identities = 151/408 (37%), Positives = 219/408 (53%), Gaps = 27/408 (6%)

Query: 451  PSKAVIDEAGQSVNGKTVLPNAELNYVAKQDFSQYKGMTASQGKIAKNFVFIDDYKDDAL 510
            P K   +E G  ++GK+VL      Y    D  QYKG  +++  I K F ++DDY ++AL
Sbjct: 1162 PHKVNKNENGVVIDGKSVLAGTTNYYELTWDLDQYKGDKSAKETIQKGFFYVDDYPEEAL 1221

Query: 511  DGKSMKVNSIKASDGTDVSQL-LEMRHVLSTDTLDEKLQTLIKEAGISPVGEFYMWTAKD 569
            D   ++ + IK +D    +   + +    S +     +Q ++K+A I+P G F ++TA D
Sbjct: 1222 D---LRTDLIKLTDANGKAVTGVSVADYASLEAAPAAVQDMLKKANITPKGAFQVFTADD 1278

Query: 570  PQAFYKAYVQKGLDVTYNLSFKVKKEFTK--GQIQNGVAQIDFGNGYTGNIVVNDLTTPE 627
            PQAFY AYV  G D+T      VK E  K  G  +N   QIDFGNGY  NIV+N++
Sbjct: 1279 PQAFYDAYVVTGTDLTIVTPMTVKAEMGKIGGSYENKAYQIDFGNGYESNIVINNVPQIN 1338

Query: 628  IHKDV---LDKEDGKSINNGTVKLGDEVTYKLEGWVVPTGRSYDLFEYKFVDQLQRTPDL 684
              KDV   +D  D  +++  T+ L     Y+L G ++P   + +LFEY F D   +T D
Sbjct: 1339 PEKDVTLTMDPADSTNVDGQTIALNQVFNYRLIGGIIPADHAEELFEYSFSDDYDQTGDQ 1398

Query: 685  YLRD-KVVAKVDVTLKDGTVIKKGTNLGEYTETVYNKKTGLYELVFKKDFLEKVARSSEF 743
            Y    K  AKVD+TLKDGT+IK GT+L  YTE   ++  G   + FK+DFL  V+  S F
Sbjct: 1399 YTGQYKAFAKVDLTLKDGTIIKAGTDLTSYTEAQVDEANGQIVVTFKEDFLRSVSVDSAF 1458

Query: 744  GADDFVVVKRIKAGDVYNTADFFINGNKVKTETVVTHTPE--KPKPVEPQ---------- 791
             A+ ++ +KRI  G   NT    +NG    + TV T TPE  +P PV+P+
Sbjct: 1459 QAEVYLQMKRIAVGTFANTYVNTVNGITYSSNTVRTSTPEPKQPSPVDPKTTTTVVFQPR 1518

Query: 792  --KATPKAPAKG--LPQTGEASVAPLTALGAIILSA-IGLAGFKKRKE 834
              KA   AP  G  LP TG++S A L  LG + L+A   L G +++++
Sbjct: 1519 QGKAYQPAPPAGAQLPATGDSSNAYLPLLGLVSLTAGFSLLGLRRKQD 1566
 Identities = 75/242 (30%), Positives = 120/242 (48%), Gaps = 33/242 (13%)

Query: 11   SADQVTTQATTQTVTQNQAETVTSTQLDKAVATAKKAAVAVTTTAAVNHATTTDAQADLA 70
            S+   T+QA     T   +  V++++LD+A  +A++A V V+  A VN   T  +  D A
Sbjct: 73   SSQAETSQAQAGQKTGAMSVDVSTSELDEAAKSAQEAGVTVSQDATVNKGTVETS--DEA 130

Query: 71   NQTQT-VKDVTAKAQANTQAIKDATAENAKIDAENKAESQRVSQLNAQTKAKID---AEN 126
            NQ +T +KD  +K  A+   I+  T +     A N+AE+ R++Q NA  KA+ +   A N
Sbjct: 131  NQKETEIKDDYSKQAAD---IQKTTEDYKAAVAANQAETDRITQENAAKKAQYEQDLAAN 187
```

```
Query: 127 KDAQAKADATNAQLQKDYQAKLAKIKSVEAYNAGVRQRNKDAQA--------------KA 172
           K    +    NAQ + DY+AKLA+ +   A    V+Q N D+QA               +
Sbjct: 188 KAEVERITNENAQAKADYEAKLAQYQKDLA---AVQQANNDSQAAYAAAKEAYDKELARV 244

Query: 173 DATNAQLQKDYQAKLA---LYNQALKAKAEADKQSINNVAFDIKAQ----AKGVDNAEYG 225
            A NA  +K+Y+  LA     N+ +KA+  A +Q       D +A+     K +  A+ G
Sbjct: 245 QAANAAAKKEYEEALAANTTKNEQIKAENAAIQQRNAQAKADYEAKLAQYEKDLAAAQSG 304

Query: 226 NS 227
           N+
Sbjct: 305 NA 306
 Identities = 63/223 (28%), Positives = 100/223 (44%), Gaps = 31/223 (13%)

Query: 2   ITTLQTSQVSADQVTTQATTQTVTQNQAETVTSTQLDKAVATAK-----------KAAVA 50
           +  +Q +  +A +   +A     T+N+     +  + + A AK           K   A
Sbjct: 241 LARVQAANAAAKKEYEEALAANTTKNEQIKAENAAIQQRNAQAKADYEAKLAQYEKDLAA 300

Query: 51  VTTTAAVNHATTTDAQADLANQTQTVKDVTAKA-QANTQAIKDATAENAKIDAENKAESQ 109
             +  A N A     +A    +   V+   A A QA  QA+   TA+NA+I AEN+A  Q
Sbjct: 301 AQSGNATNEADYQAKKAAYEQELARVQAANAAAKQAYEQALAANTAKNAQITAENEAIQQ 360

Query: 110 RVSQLNAQTKAKIDAENKDAQAKADATNAQLQKDYQAKLA----KIKSVEAYNAGVRQRN 165
           R +Q  A  +AK+    KD  A A + NA  + DYQ KLA    ++  V+A NA  +Q
Sbjct: 361 RNAQAKANYEAKLAQYQKDL-AAAQSGNAANEADYQEKLAAYEKELARVQAANAAAKQEY 419

Query: 166 KDAQAKADATNAQL--------------QKDYQAKLALYNQAL 194
           +    +A+A NA++              + DY+ KL+ Y + L
Sbjct: 420 EQKVQEANAKNAEITEANRAIRERNAKAKTDYELKLSKYQEEL 462
 Identities = 75/243 (30%), Positives = 101/243 (40%), Gaps = 56/243 (23%)

Query: 8   SQVSAD-QVTTQATTQTVTQNQAETVTSTQLDKAVATAKKAAVAVTTTAAVNHATTTDAQ 66
           S+ +AD Q TT+     V  NQAET   TQ + A   A+         A V   T  +AQ
Sbjct: 142 SKQAADIQKTTEDYKAAVAANQAETDRITQ-ENAAKKAQYEQDLAANKAEVERITNENAQ 200

Query: 67  ADL---ANQTQTVKDVTAKAQANT--------------------------------QAIK 91
           A     A   Q  KD+ A  QAN                                 +A+
Sbjct: 201 AKADYEAKLAQYQKDLAAVQQANNDSQAAYAAAKEAYDKELARVQAANAAAKKEYEEALA 260

Query: 92  DATAENAKIDAENKAESQRVSQLNAQTKAKIDAENKDAQAKADATNAQLQKDYQAKLA-- 149
             T +N +I AEN A  QR +Q  A  +AK+    KD  A A + NA  + DYQAK A
Sbjct: 261 ANTTKNEQIKAENAAIQQRNAQAKADYEAKLAQYEKDL-AAAQSGNATNEADYQAKKAAY 319

Query: 150 --KIKSVEAYNAGVRQRNKDAQAKADATNAQL--------------QKDYQAKLALYNQA 193
             ++  V+A NA  +Q  + A A   A NAQ+              + +Y+AKLA Y +
Sbjct: 320 EQELARVQAANAAAKQAYEQALAANTAKNAQITAENEAIQQRNAQAKANYEAKLAQYQKD 379

Query: 194 LKA 196
           L A
Sbjct: 380 LAA 382
```

There is also homology to SEQ ID 598.

SEQ ID 1814 (GBS191) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 176 (lane 2; MW 91kDa).

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 582

A DNA sequence (GBSx0622) was identified in *S.agalactiae* <SEQ ID 1815> which encodes the amino acid sequence <SEQ ID 1816>. This protein is predicted to be TnpA. Analysis of this protein sequence reveals the following:

```
Possible site: 34
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2935(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10221> which encodes amino acid sequence <SEQ ID 10222> was also identified.

A related GBS nucleic acid sequence <SEQ ID 9921> which encodes amino acid sequence <SEQ ID 9922> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAC82523 GB:AF027768 TnpA [Serratia marcescens]
 Identities = 168/385 (43%), Positives = 232/385 (59%), Gaps = 13/385 (3%)

Query: 26  MMFKVEAVGPPERCPECGFD-KLYKHSSRNQLIMDLPIRLKRVGLHLNRRRYKCRECGST 84
           M F+V+ V  P  C ECG   +  +   R+    DLPI  KRV L + RRRY CR C +T
Sbjct: 1   MHFQVD-VPDPIACEECGVQGEFVRFGKRDVPYRDLPIHGKRVTLWVVRRRYTCRACKTT 59

Query: 85  IS------VDEKRSMTKRLLKSIQEQSMSKTFVEVAESVGVDEKTIRNVFKDYVALKERE 138
                   VD  R MT RL + ++++S +  +  VA   G+DEKT+R++F        R
Sbjct: 60  FRPQLPEMVDGFR-MTLRLHEYVEKESFNHPYTFVAAQTGLDEKTVRDIFNARAEFLGRW 118

Query: 139 YQFETPKWLGIDEIHIIRRPRLVLTNIERRTIYDIKPNRNKETVIQRLSEISDRTYIEYV 198
           ++FETP+ LGIDE+++ +R R +LTNIE RT+ D+   R ++ V   L ++ DR  +E V
Sbjct: 119 HRFETPRILGIDELYLNKRYRCILTNIEERTLLDLLATRRQDVVTNYLMKLKDRQKVEIV 178

Query: 199 TMDMWKPYKDAVNTILPQAKVVVDKFHVVRMANQALDNVRKSLKAHMSQKERRTLMRERF 258
           +MDMW PY+ AV  +LPQA++VVDKFHVVRMAN AL+ VRK L+  +   + RTL  +R
Sbjct: 179 SMDMWNPYRAAVKAVLPQARIVVDKFHVVRMANDALERVRKGLRKELKPSQSRTLKGDRK 238

Query: 259 ILLKRKHDLNERESFLLDTWLGNLPALKEAYELKEEFYWIWDTPDPDEGHLRYSQWRHRC 318
           ILLKR H++++RE   +++TW G  P L  AYE KE FY IWD     +      +W
Sbjct: 239 ILLKRAHEVSDRERLIMETWTGAFPQLLAAYEHKERFYGIWDATTRLQAEAALDEWI-AT 297

Query: 319 MSSNSKDAYKDLVRAVDNWHVEIFNYF--DKRLTNAYTESINSIIRQVERMGRGYSFDAL 376
           +    K+ + DLVRAV NW  E   YF  D  +TNAYTESIN + +   R GRGYSF+ +
Sbjct: 298 IPKGQKEVWSDLVRAVGNWREETMTYFETDMPVTNAYTESINRLAKDKNREGRGYSFEVM 357

Query: 377 RAKILFNEKLHKKRKPRFNSSAFNK 401
           RA++L+  K HKK+ P    S F K
Sbjct: 358 RARMLYTTK-HKKKAPTAKVSPFYK 381
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 583

A DNA sequence (GBSx0623) was identified in *S.agalactiae* <SEQ ID 1817> which encodes the amino acid sequence <SEQ ID 1818>. This protein is predicted to be mercuric reductase. Analysis of this protein sequence reveals the following:

```
Possible site: 53
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2115(Affirmative) < succ>
```

```
                  bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                   bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAA70224 GB:Y09024 mercuric reductase [Bacillus cereus]
 Identities = 411/546 (75%), Positives = 483/546 (88%)

Query: 1   MNKFKVNISGMTCTGCEKHVESALEKIGAKNIESSYRRGEAVFELPDDIEVESAIKAIDE 60
           M K++V++ GMTCTGCE+HV  ALE +GA  IE  +RRGEAVFELP+ + VE+A KAI +
Sbjct: 1   MKKYRVDVQGMTCTGCEEHVAVALENMGATGIEVDFRRGEAVFELPNALGVETAKKAISD 60

Query: 61  ANYQAGEIEEVSSLENVALINEDNYDLLIIGSGAAAFSSAIKAIEYGAKVGMIERGTVGG 120
           A YQ G+ EEV S E V L NE +YD +IIGSG AAFSSAI+A++YGAKV MIERGT+GG
Sbjct: 61  AKYQPGKAEEVQSQEMVQLGNEGDYDYIIIGSGGAAFSSAIEAVKYGAKVAMIERGTIGG 120

Query: 121 TCVNIGCVPSKTLLRAGEINHLSKDNPFIGLQTSAGEVDLASLITQKDKLVSELRNQKYM 180
           TCVNIGCVPSKTLLRAGEINHL+K+NPF+GL TSAGEVDLA LI QK++LV+ELRN KY+
Sbjct: 121 TCVNIGCVPSKTLLRAGEINHLAKNNPFVGLHTSAGEVDLAPLIKQKNELVTELRNSKYV 180

Query: 181 DLIDEYNFDLIKGEAKFVDASTVEVNGTKLSAKRFLIATGASPSLPQISGLEKMDYLTST 240
           DLID+Y F+LI+GEAKFVD  TVEVNG  +SAKRFLIATGASP+ P I GL ++DYLTST
Sbjct: 181 DLIDDYGFELIEGEAKFVDEKTVEVNGAPISAKRFLIATGASPAKPNIPGLNEVDYLTST 240

Query: 241 TLLELKKIPKRLTVIGSGYIGMELGQLFHHLGSEITLMQRSERLLKEYDPEISESVEKAL 300
           +LLELKK+PKRL VIGSGYIGMELGQLFH+LGSE+TL+QRSERLLKEYDPEISESVEK+L
Sbjct: 241 SLLELKKVPKRLVVIGSGYIGMELGQLFHNLGSEVTLIQRSERLLKEYDPEISESVEKSL 300

Query: 301 IEQGINLVKGATFERVEQSGEIKRVYVTVNGSREVIESDQLLVATGRKPNTDSLNLSAAG 360
           +EQGINLVKGAT+ER+EQ+G+IK+V+V VNG + +IE+DQLLVATGR PNT +LNL AAG
Sbjct: 301 VEQGINLVKGATYERIEQNGDIKKVHVEVNGKKRIIEADQLLVATGRTPNTATLNLRAAG 360

Query: 361 VETGKNNEILINDFGQTSNEKIYAAGDVTLGPQFVYVAAYEGGIITDNAIGGLNKKIDLS 420
           VE G   EI+I+D+ +T+N +IYAAGDVTLGPQFVYVAAY+GG+   NAIGGLNKK++L
Sbjct: 361 VEIGSRGEIIIDDYSRTTNTRIYAAGDVTLGPQFVYVAAYQGGVAAPNAIGGLNKKLNLE 420

Query: 421 VVPAVTFTNPTVATVGLTEEQAKEKGYDVKTSVLPLGAVPRAIVNRETTGVFKLVADAET 480
           VVP VTFT P +ATVGLTE+QAKE GY+VKTSVLPL AVPRA+VNRETTGVFKLVAD++T
Sbjct: 421 VVPGVTFTAPAIATVGLTEQQAKENGYEVKTSVLPLDAVPRALVNRETTGVFKLVADSKT 480

Query: 481 LKVLGVHIVSENAGDVIYAASLAVKFGLTIEDLTETLAPYLTMAEGLKLVALTFDKDISK 540
           +KVLG H+V+ENAGDVIYAA+LAVKFGLT++D+ ETLAPYLTMAEGLKL ALTFDKDISK
Sbjct: 481 MKVLGAHVVAENAGDVIYAATLAVKFGLTVDDIRETLAPYLTMAEGLKLAALTFDKDISK 540

Query: 541 LSCCAG 546
           LSCCAG
Sbjct: 541 LSCCAG 546
```

There is also homology to SEQ ID 1820.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 584

A DNA sequence (GBSx0624) was identified in *S.agalactiae* <SEQ ID 1821> which encodes the amino acid sequence <SEQ ID 1822>. This protein is predicted to be regulatory protein. Analysis of this protein sequence reveals the following:

```
Possible site: 36
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.4529(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
```

```
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAA83973 GB:AF138877 mercury resistance operon negative
          regulator MerR1 [Bacillus sp. RC607]
 Identities = 84/129 (65%), Positives = 105/129 (81%)

Query: 1   MIYRISEFADKCGVNKETIRYYERKNLLQEPHRTEAGYRIYSYDDVKRVGFIKRIQELGF 60
            M +RI E ADKCGVNKETIRYYER  L+ EP RTE GYR+YS   V R+ FIKR+QELGF
Sbjct: 1   MKFRIGELADKCGVNKETIRYYERLGLIPEPERTEKGYRMYSQQTVDRLHFIKRMQELGF 60

Query: 61  SLSEIYKLLGVVDKDEVRCQDMFEFVSKKQKEVQKQIEDLKRIETMLDDLKQRCPDEKKL 120
            +L+EI KLLGVVD+DE +C+DM++F   K +++Q++IEDLKRIE ML DLK+RCP+ K +
Sbjct: 61  TLNEIDKLLGVVDRDEAKCRDMYDFTILKIEDIQRKIEDLKRIERMLMDLKERCPENKDI 120

Query: 121 HSCPIIETL 129
            + CPIIETL
Sbjct: 121 YECPIIETL 129
```

There is also homology to SEQ ID 1712.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 585

A DNA sequence (GBSx0625) was identified in *S.agalactiae* <SEQ ID 1823> which encodes the amino acid sequence <SEQ ID 1824>. This protein is predicted to be Nramp metal ion transporter. Analysis of this protein sequence reveals the following:

```
Possible site: 53
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-13.85    Transmembrane  175 - 191 ( 169 - 201)
   INTEGRAL    Likelihood =-11.94    Transmembrane  150 - 166 ( 132 - 173)
   INTEGRAL    Likelihood = -9.45    Transmembrane  491 - 507 ( 481 - 509)
   INTEGRAL    Likelihood = -8.92    Transmembrane  375 - 391 ( 374 - 396)
   INTEGRAL    Likelihood = -8.39    Transmembrane   72 -  88 (  69 -  93)
   INTEGRAL    Likelihood = -7.96    Transmembrane  280 - 296 ( 274 - 299)
   INTEGRAL    Likelihood = -7.17    Transmembrane  413 - 429 ( 411 - 431)
   INTEGRAL    Likelihood = -6.79    Transmembrane  327 - 343 ( 322 - 346)
   INTEGRAL    Likelihood = -3.40    Transmembrane  444 - 460 ( 443 - 462)
   INTEGRAL    Likelihood = -3.24    Transmembrane  132 - 148 ( 132 - 149)
   INTEGRAL    Likelihood = -0.96    Transmembrane  115 - 131 ( 114 - 131)

----- Final Results -----
               bacterial membrane --- Certainty=0.6540(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAF83825 GB:AE003939 manganese transport protein [Xylella
          fastidiosa]
 Identities = 185/450 (41%), Positives = 278/450 (61%), Gaps = 29/450 (6%)

Query: 16  ANGPSLEEINGTIEVPKDLSFFKTLLAYSGPGALVAVGYMDPGNWSTSITGGQNFQYLLI 75
            ++ PSL E++ ++ V +   +   LLA+ GPG +V+VGYMDPGNW+T + GG  F Y+L+
Sbjct: 35  SDSPSLGEMHASVAVSRRGHWGFRLLAFLGPGYMVSVGYMDPGNWATGLAGGSRFGYMLL 94

Query: 76  SIILMSSLIAMLLQYMSAKLGIVTQMDLAQAIRARTSKQLGIVLWILTELAIMATDIAEV 135
            S+IL+S+++A++LQ ++A+LGI + MDLAQA RAR S+   + LW++ ELAI+A D+AEV
Sbjct: 95  SVILLSNVMAIVLQALAARLGIASDMDLAQACRARYSRGTTLALWVVCELAIIACDLAEV 154

Query: 136 IGGAIALYLLFHIPLAIAVFITVFDVLLLLLLTKIGFRKIEALVVALILVIFLVFAYQVA 195
```

```
           IG AIAL LL  +P+   V IT  DV+L+LLL   GFR +EA V+AL+LVIF  F  Q+
Sbjct: 155 IGTAIALNLLLGVPIIWGVVITAVDVVLVLLLMHRGFRALEAFVIALLLVIFGCFVVQIV 214

Query: 196 LSHPIWTDIFKGLVPTSEAFSTSHTVNGQTPLSGALGIIGATVMPHNLYLHSSVVQSRKL 255
           L+ P    ++ G VP +        V     L  A+GI+GATVMPHNLYLHSS+VQ+R
Sbjct: 215 LAAPPLQEVLGGFVPRWQ------VVADPQALYLAIGIVGATVMPHNLYLHSSIVQTRAY 268

Query: 256 DHNNKKDIAR--AIRFSTFDSNIQLTVAFFVNSLLLIMGVAVFKTGSVTDPSFFGLFKAL 313
               +  + R  A+R++  DS + L +A F+N+ +LI+  AVF      D
Sbjct: 269 P---RTPVGRRSALRWAVADSTLALMLALFINASILILAAAVFHAQHHFD---------- 315

Query: 314 SNSTIMSNSILAHIASSGILSLLFAIALLASGQNSTITGTLTGQIIMEGFIHMKVPIWFR 373
                 +  +LA +   G+ + LFA ALLASG NST+T TL GQI+MEGF+ +++  W R
Sbjct: 316 VEEIEQAYQLLAPVLGVGVAATLFATALLASGINSTVTATLAGQIVMEGFLRLRLRPWLR 375

Query: 374 RIITRLISVIPVMICVLVTSGRSTVEEHIAINNLMNNSQVFLAFALPFSMLPLLIFTNSK 433
           R++TR ++++PV++ V +   + T         L+  SQV L+  LPF+++PLL     +
Sbjct: 376 RVLTRGLAIVPVIVVVALYGEQGT-------GRLLLLSQVILSMQLPFAVIPLLRCVADR 428

Query: 434 VEMDDDFKNTWIIKILGWLSVIGLIYLNMK 463
             M       W++ ++ WL    ++ LN+K
Sbjct: 429 KVMGALVAPRWLM-VVAWLIAGVIVVLNVK 457
```

No corresponding DNA sequence was identified in *S.pyogenes.*

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 586

A DNA sequence (GBSx0626) was identified in *S.agalactiae* <SEQ ID 1825> which encodes the amino acid sequence <SEQ ID 1826>. Analysis of this protein sequence reveals the following:

```
Possible site: 20
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2590(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes.*

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 587

A DNA sequence (GBSx0627) was identified in *S.agalactiae* <SEQ ID 1827> which encodes the amino acid sequence <SEQ ID 1828>. Analysis of this protein sequence reveals the following:

```
Possible site: 53
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -9.82   Transmembrane  212 - 228 ( 204 - 233)
   INTEGRAL    Likelihood = -8.39   Transmembrane   98 - 114 (  94 - 125)
   INTEGRAL    Likelihood = -7.22   Transmembrane  132 - 148 ( 122 - 154)
   INTEGRAL    Likelihood = -6.42   Transmembrane  159 - 175 ( 155 - 188)
   INTEGRAL    Likelihood = -4.78   Transmembrane   54 -  70 (  51 -  72)
   INTEGRAL    Likelihood = -2.97   Transmembrane   18 -  34 (  15 -  36)

----- Final Results -----
```

```
     bacterial membrane --- Certainty=0.4927(Affirmative) < succ>
      bacterial outside --- Certainty=0.0000(Not Clear) < succ>
    bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB16051 GB:Z99124 yydJ [Bacillus subtilis]
 Identities = 97/239 (40%), Positives = 154/239 (63%), Gaps = 3/239 (1%)

Query: 4   LEFRKSIRGRTLFYIISTVALTYVLGYILPVGIDKIRHLTLGEFYFSTYTVFTQFGFLIF 63
            LEF+KSI  + +  + +     ++LGY L VGIDK+ ++T   F+FS+YTV TQFG ++F
Sbjct: 3   LEFKKSISNKVIIILGAMFVFLFLLGYFLLVGIDKVSNVTPEMFFFSSYTVATQFGLMLF 62

Query: 64  GFVIVYFFNKDYSDKCILYHYFSGYHLTKYFYTKLLVLFSEFFIAIIVCNILASLLWGYS 123
             FVI +F N++YS+K IL++   G ++  +FY K+ VLF E F  I +  ++ SL++ +
Sbjct: 63  SFVIAFFINREYSNKNILFYKLIGENIYTFFYKKIAVLFLECFAFITLGLLIISLMY-HD 121

Query: 124 LFYFLTTTILFSLVVLQYLLVVSTISILFSNMLVSIGVTIFYWITSIILVAIGG-IFKVS 182
              +F     LFS V+LQY+L++ TIS+L  N+L+SIGV+I YW+TS+ILVAI    F
Sbjct: 122 FSHFALLLFLFSAVILQYILIIGTISVLCPNILISIGVSIVYWMTSVILVAISNKTFGFI 181

Query: 183 AIFDASNSLYKIIGK-LFSHPMTIDLTDFFIIVPYMICLSVISFLIVCLSNRRWLLNGM 240
            A F+A N++Y  I + L S  MT+   D   I+ Y++ + +I+ +++  S  RW+  G+
Sbjct: 182 APFEAGNTMYPRIERVLQSDNMTLGSNDVLFIILYLVSIIIINAIVLRFSKTRWIKMGL 240
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 588

A DNA sequence (GBSx0628) was identified in *S.agalactiae* <SEQ ID 1829> which encodes the amino acid sequence <SEQ ID 1830>. This protein is predicted to be antibiotic epidermin immunity protein F. Analysis of this protein sequence reveals the following:

```
Possible site: 43
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
           bacterial cytoplasm --- Certainty=0.2901(Affirmative) < succ>
            bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
             bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB16052 GB:Z99124 similar to ABC transporter (ATP-binding
           protein) [Bacillus subtilis]
 Identities = 100/209 (47%), Positives = 150/209 (70%), Gaps = 4/209 (1%)

Query: 1   MFINNYTLKIGNRILLENTNLDFEEGEINHLLGRNGSGKSQLAKDFIINRGNYFSNDIYE 60
            M I NYTLK+  + LL++T+L F  G+INH++G+NG GKSQLAKDF++N       DI +
Sbjct: 1   MNIANYTLKVKGKTLLQDTDLHFSSGKINHVVGKNGVGKSQLAKDFLLNNSKRIGRDIRQ 60

Query: 61  DTLIISSYSNLPSDVT----INDLERTIPWKLSKEIYQLLNINQISKTVKLKQLSDGQKQ 116
            +  +ISS SN+P+DV+    ++ L +    K+  +I  LLN++ I   V +K LSDGQKQ
Sbjct: 61  NVSLISSSSNIPNDVSKDFLLHFLSKKFDAKMIDKIAYLLNLDNIDGKVLIKNLSDGQKQ 120

Query: 117 KVKLLVLLSLDKHIIILDEITNALDKKSVDEINVFLQNYIQYYPEKIIINISHDINNIRS 176
            K+KLL  L  DK+II+LDEITN+LDKK+V EI+ FL  YIQ  PEKIIINI+HD++++++
Sbjct: 121 KLKLLSFLLEDKNIIVLDEITNSLDKKTVIEIHGFLNKYIQENPEKIIINITHDLSDLKA 180

Query: 177 LKGNYFLIDNQKICKVDTLDDAISWYLGE 205
            ++G+Y++ ++Q+I +  ++D  I  Y+ E
Sbjct: 181 IEGDYYIFNHQEIQQYHSVDKLIEVYINE 209
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1831> which encodes the amino acid sequence <SEQ ID 1832>. Analysis of this protein sequence reveals the following:

```
Possible site: 13
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2760(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 49/174 (28%), Positives = 82/174 (46%), Gaps = 27/174 (15%)

Query: 3   INNYTLKIGNRILLENTNLDFEEGEINHLLGRNGSGKSQLAK----------DFIINRGN 52
           I N     G R +L N N++  +G++  L+G NG+GKS + K            II  G
Sbjct: 23  IQNLKKSYGKRTILNNVNMNIPKGKVYALIGPNGAGKSTIMKILTGLVSKTSGSIIFEGR 82

Query: 53  YFS-------NDIYEDTLI---ISSYSNLPSDVTINDL-ERTIPWKLSKEIYQLLNINQI 101
            +S         I E+  +   +S+Y N+    T+  + E TI   L+K     + +  I
Sbjct: 83  EWSRRDLRKIGSIIEEPPLYKNLSAYDNMKVVTTMLGVSESTILPLLNK-----VGLGNI 137

Query: 102 SKTVKLKQLSDGQKQKVKLLVLLSLDKHIIILDEITNALDKKSVDEINVFLQNY 155
            K   +KQ S G KQ++ + + L     ++ILDE TN LD   + E+   ++++
Sbjct: 138 DKR-PVKQFSLGMKQRLGIAISLINSPKLLILDEPTNGLDPIGIQELREIIESF 190
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 589

A DNA sequence (GBSx0629) was identified in *S.agalactiae* <SEQ ID 1833> which encodes the amino acid sequence <SEQ ID 1834>. This protein is predicted to be aminoglycoside 6-adenylyltransferase. Analysis of this protein sequence reveals the following:

```
Possible site: 33
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1780(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAA29839 GB:X06627 ORF (str) [Staphylococcus aureus]
 Identities = 91/289 (31%), Positives = 146/289 (50%), Gaps = 14/289 (4%)

Query: 1   MRDEQEIYNLVLNIANQDKRIEAVLLNGSRANPNVPKDDFQDYDIVFVTNFIEDIISDTN 60
           MR E+EI NLV   A Q   ++ + L GSR N N+ KD FQDYD  F  + IE    + +
Sbjct: 1   MRTEKEILNLVSEFAYQRSNVKIIALEGSRTNENIKKDKFQDYDFAFFVSDIEYFTHEES 60

Query: 61  YHKKFGDILIMQKPNE---FRNKTEYNCFAYLMQFQDLTRIDLRLIKPEFLEDYLDDA-- 115
           +   FG++L +QKP +   F    +Y  ++Y+M F+D  ++D+ LI  + L  Y  D+
Sbjct: 61  WLSLFGELLFIQKPEDMELFPPDLDYG-YSYIMYFKDGIKMDITLINLKDLNRYFSDSDG 119

Query: 116 FSKVLLDKKNKYLDYNFERSSLYETKQLSEDEINKILNEIYWVSTYVVKGIARNDIIYSE 175
             K+L+DK N          S Y  K+ +E E     NE + VSTYV KG+ R +I+++
Sbjct: 120 LVKILVDKDNLVTQEIVPDDSNYWLKKPTEREFYDCCNEFWSVSTYVAKGVFRREILFAL 179

Query: 176 FMISNPIKNAFIKLLKQKILIEKELDSLSFGKLDKDILQYITDKD--QLLKIFSNKSLKD 233
              +N ++   ++++   I   +  D  S GK  K I +Y+TDK+   LL  F     +
```

```
Sbjct: 180 DHFNNILRPELLRMISWYIGFNRGFD-FSLGKNYKFINKYLTDKEFNMLLATFEMNGYRK 238

Query: 234 IEANLRFLLDETNQMAKYISINRKLNLNQGEYQSAMKFMNIFLSNSYQN 282
              + +       ++ KY S N+   L    Y +  K +  F+ N+Y+N
Sbjct: 239 TYQSFKLCC----ELFKYYS-NKVSCLGNYNYPNYEKNIENFIRNNYEN 282
```

No corresponding DNA sequence was identified in *S.pyogenes*.

A related GBS gene <SEQ ID 8607> and protein <SEQ ID 8608> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 5
McG: Discrim Score:      -5.26
GvH: Signal Score (-7.5): -6.14
     Possible site: 33
>>> Seems to have no N-terminal signal sequence
ALOM program   count: 0 value:   6.10 threshold:  0.0
   PERIPHERAL  Likelihood =  6.10     151
 modified ALOM score:  -1.72

*** Reasoning Step: 3

----- Final Results -----
               bacterial cytoplasm --- Certainty=0.1780(Affirmative) < succ>
                bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                 bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
31.0/53.4% over 281aa
 Staphylococcus aureus
  EGAD|9462| streptomycin resistance protein Insert characterized
  SP|P12055|STR_STAAU STREPTOMYCIN RESISTANCE PROTEIN. Insert characterized
  GP|46644|emb|CAA29839.1||X06627 ORF (str) Insert characterized
  PIR|S00938|S00938 str protein - plasmid pS194 Insert characterized

ORF00399(301 - 1146 of 1452)
EGAD|9462|9267(1 - 282 of 282) streptomycin resistance protein {Staphylococcus aureus}
SP|P12055|STR_STAAU STREPTOMYCIN RESISTANCE PROTEIN. GP|46644|emb|CAA29839.1||X06627 ORF
(str) {Staphylococcus aureus} PIR|S00938|S00938 str protein - Staphylococcus aureus plasmid
pS194
%Match = 12.8
%Identity = 31.0  %Similarity = 53.4
Matches = 87  Mismatches = 125  Conservative Sub.s = 63

117       147       177       207       237       267       297       327
**LMTY*H*TVENIWNHNQLLRKI*N*ILGGRKGMSMLI*VYDYMLREKYKGNIKVLEXTW*YKVK*EVAIMRDEQEIYN
                                                                       || |:|| |
                                                                       MRTEKEILN

357       387       417       447       477       507                 558
LVLNIANQDKRIEAVLLNGSRANPNVPKDDFQDYDIVFVTNFIEDIISDTNYHKKFGDILIMQKPNEFR---NKTEYNCF
||   | |   :: : | ||| | |: || |||||  |  : ||    : ::   ||::| :||| :        :|  :
LVSEFAYQRSNVKIIALEGSRTNENIKKDKFQDYDFAFFVSDIEYFTHEESWLSLFGELLFIQKPEDMELFPPDLDYG-Y
         20        30        40        50        60        70        80

588       618                 672       702       732       762       792
AYLMQFQDLTRIDLRLIKPEFLEDYLDDA--FSKVLLDKKNKYLDYNFERSSLYETKQLSEDEINKILNEIYWVSTYVVK
:|:| |:|  ::|: ||  : |  |: |:  : |:|:|| |          | |  |: :| |     || : ||||| |
SYIMYFKDGIKMDITLINLKDLNRYFSDSDGLVKILVDKDNLVTQEIVPDDSNYWLKKPTEREFYDCCNEFWSVSTYVAK
         100       110       120       130       140       150       160

822       852       882       912       942       966       996       1026
GIARNDIIYSEFMISNPIKNAFIKLLKQKILIEKELDSLSFGKLDKDILQYITDKD--QLLKIFSNKSLKDIEANLRFLL
|: |:|:::    :| ::  :::::   |   : :| :|:||  | | :|:|||:   ||  |     :    ::::
GVFRREILFALDHFNNILRPELLRMISWYIGFNRGFD-FSLGKNYKFINKYLTDKEFNMLLATFEMNGYRKTYQSFKLCC
         180       190       200        210       220       230       240
```

```
1056      1086      1116      1146      1176      1206      1236      1266
DETNQMAKYISINRKLNLNQGEYQSAMKFMNIFLSNSYQNFN*YYCVKDNRL*LSKLNYHS*RFSRKIINNFGDK*WDKS
:     :  :|      |    | :  | :  |: |:|:|
ELFKYYSNKVS-----CLGNYNYPNYEKNIENFIRNNYEN
               260       270       280
```

SEQ ID 1834 (GBS46) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 11 (lane 6; MW 34.9kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 16 (lane 3; MW 59.8kDa).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 590

A DNA sequence (GBSx0630) was identified in *S.agalactiae* <SEQ ID 1835> which encodes the amino acid sequence <SEQ ID 1836>. Analysis of this protein sequence reveals the following:

```
Possible site: 29
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.1179(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 591

A DNA sequence (GBSx0631) was identified in *S.agalactiae* <SEQ ID 1837> which encodes the amino acid sequence <SEQ ID 1838>. Analysis of this protein sequence reveals the following:

```
Possible site: 44
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -2.81   Transmembrane  177 - 193 ( 177 - 194)
   INTEGRAL    Likelihood = -0.27   Transmembrane  129 - 145 ( 129 - 145)

----- Final Results -----
              bacterial membrane --- Certainty=0.2126(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 8609> which encodes amino acid sequence <SEQ ID 8610> was also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 9
McG: Discrim Score:    -19.59
GvH: Signal Score (-7.5): -4.49
     Possible site: 44
>>> Seems to have no N-terminal signal sequence
ALOM program   count: 2 value:  -2.81 threshold:  0.0
   INTEGRAL    Likelihood = -2.81   Transmembrane  172 - 188 ( 172 - 189)
   INTEGRAL    Likelihood = -0.27   Transmembrane  124 - 140 ( 124 - 140)
   PERIPHERAL  Likelihood =  8.01      30
```

```
 modified ALOM score:   1.06

*** Reasoning Step: 3

----- Final Results -----
              bacterial membrane --- Certainty=0.2126(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 592

A DNA sequence (GBSx0632) was identified in *S.agalactiae* <SEQ ID 1839> which encodes the amino acid sequence <SEQ ID 1840>. Analysis of this protein sequence reveals the following:

```
Possible site: 27
>>> Seems to have an uncleavable N-term signal seq

----- Final Results -----
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10223> which encodes amino acid sequence <SEQ ID 10224> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB49414 GB:AJ248284 hypothetical protein [Pyrococcus abyssi]
 Identities = 29/86 (33%), Positives = 52/86 (59%), Gaps = 4/86 (4%)

Query: 14  TYYILLALFE--EAHGYAIMQKVEEMSGGDVRIAAGTMYGAIENLLKQKWIKSIPSD--D 69
             +Y ILL L E  + HGYAI +++EE++ G +  + G +Y  ++ L K K ++    ++
Sbjct: 19  SYLILLILNENEKLHGYAIRKRLEELTDGKLVPSEGALYSILKMLKKYKLVEDYWAEVGG 78

Query: 70  RRRKVYIITETGKEIVELETNRLRKL 95
            R R+ Y ITE GKE+++     +R++
Sbjct: 79  RVRRYYQITELGKEVLDEIKEEIREI 104
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 593

A DNA sequence (GBSx0633) was identified in *S.agalactiae* <SEQ ID 1841> which encodes the amino acid sequence <SEQ ID 1842>. Analysis of this protein sequence reveals the following:

```
Possible site: 23
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.0510(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10225> which encodes amino acid sequence <SEQ ID 10226> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAF22299 GB:AF185571 putative N-acetyltransferase Camello 2
          [Homo sapiens]
 Identities = 32/110 (29%), Positives = 54/110 (49%), Gaps = 4/110 (3%)

Query: 67  IKMAEQDDIFQIENYYQNRKGQ-FWIALENEKVVGSIALLRIDDKTAVLKKFFTYPKYRG 125
           + +A + D+  I   Y +  G  FW+A   EKVVG++  L +DD T   K+   +
Sbjct: 86  VDIALRTDMSDITKSYLSECGSCFWVAESEEKVVGTVGALPVDDPTLREKRLQLFHLSVD 145

Query: 126 NPVR---LGRKLFERFMLFARASKFTRIVLDTPEKEKRSHFFYENQGFKQ 172
           N  R   + + L    + FAR   ++ +VLDT   +  +   Y++ GFK+
Sbjct: 146 NEHRGQGIAKALVRTVLQFARDQGYSEVVLDTSNIQLSAMGLYQSLGFKK 195
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 594

A DNA sequence (GBSx0634) was identified in *S.agalactiae* <SEQ ID 1843> which encodes the amino acid sequence <SEQ ID 1844>. Analysis of this protein sequence reveals the following:

```
Possible site: 47
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood =-11.94   Transmembrane  159 - 175 ( 151 - 180)
   INTEGRAL    Likelihood =-11.62   Transmembrane  231 - 247 ( 225 - 251)
   INTEGRAL    Likelihood = -9.98   Transmembrane  182 - 198 ( 177 - 203)
   INTEGRAL    Likelihood = -7.11   Transmembrane  118 - 134 ( 106 - 136)
   INTEGRAL    Likelihood = -1.49   Transmembrane   74 -  90 (  74 -  93)

----- Final Results -----
               bacterial membrane --- Certainty=0.5776(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10227> which encodes amino acid sequence <SEQ ID 10228> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB15891 GB:Z99123 yxlG [Bacillus subtilis]
 Identities = 42/188 (22%), Positives = 94/188 (49%), Gaps = 4/188 (2%)

Query: 1   MKSLAVMLKKEWMENVRTYKVISILITCSIFGILGPLTALMMPDIMA--GILPKKLQGAI 58
           MK +  +L+KEW+E  ++ K+I + I   I G+  PLT   MP+I+A  G LP  ++ +
Sbjct: 1   MKVMMALLQKEWLEGWKSGKLIWLPIAMMIVGLTQPLTIYYMPEIIAHGGNLPDGMKISF 60

Query: 59  PEPTYIDSYIQYFKNMNQLGLVILVFLFSSTLTQEFSKGTLINLVTKGLAKKVIILAKFI 118
             P+  +  +      N LG+ +++F    ++  E ++G    ++++ +     I++K++
Sbjct: 61  TMPSGSEVMVSTLSQFNTLGMALVIFSVMGSVANERNQGVTALIMSRPVTAAHYIVSKWL 120

Query: 119 VITLLWTVSYLLSVVIHFSYTLYYFSNEGSHKLMVYGATWFIGILFI-SLILFFSVLFRK 177
           + +++  +S+     + + Y    F +    +       + + ++FI +  L  S +FR
Sbjct: 121 IQSVIGIMSFAAGYGLAYYYVRLLFEDASFSRFAASLGLYALWVIFIVTAGLAGSTIFR- 179

Query: 178 TLGGLLGC 185
           ++G    C
Sbjct: 180 SVGAAAAC 187
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 595

A DNA sequence (GBSx0635) was identified in *S.agalactiae* <SEQ ID 1845> which encodes the amino acid sequence <SEQ ID 1846>. This protein is predicted to be ABC transporter, ATP-binding protein. Analysis of this protein sequence reveals the following:

```
Possible site: 14
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3431(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 10229> which encodes amino acid sequence <SEQ ID 10230> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB12736 GB:Z99108 similar to ABC transporter (ATP-binding
           protein) [Bacillus subtilis]
 Identities = 105/299 (35%), Positives = 175/299 (58%), Gaps = 11/299 (3%)

Query: 4   ISFQNVTKSFGPKKILNNVSFDLEENMIYGFVGPNGAGKTTTIKMILGLLKFDTGFITIF 63
           +  +NVTK+   + I++++SF + E  ++GF+GPNGAGKTTTI+M++GL+K   G + I
Sbjct: 5   LELKNVTKNIRGRTIIDDLSFTIREGEVFGFLGPNGAGKTTTIRMMVGLMKLSKGDVLIC 64

Query: 64  GKKVNFGRTDTNQLIGYLPDVPEYYDYMTALEYLDLCSGLARSKHKLSNKELLRSVGLDD 123
           G+ +        + IG + + PE Y +++  + L   + + +   K    E++  VGL D
Sbjct: 65  GQSITKEYAKAIKHIGAIVENPELYKFLSGYKNLQQFARMVKGVTKEKIDEVVELVGLTD 124

Query: 124 N-HQKIATYSRGMKQRLGLAQALVHDPKIIICDEPTSALDPKGRQDILDIISNLRGEK-- 180
             H K+ TYS GM+QRLGLAQ L+HDPK++I DEPT+ LDP G ++I D +  L  E+
Sbjct: 125 RIHDKVKTYSLGMRQRLGLAQCLLHDPKVLILDEPTNGLDPAGIREIRDHLKKLTRERGM 184

Query: 181 TVIFSTHILSDVEKICDHVLVLTKCGIYSLEELKGKKSEENYSVRILIKVTKSEAKVLSH 240
            VI S+H+LS++E +CD + +L K  +  ++ +K +  +EN +    ++   SEA  + +
Sbjct: 185 AVIVSSHLLSEMELMCDRIAILQKGKLIDIQNVKDENIDENDTYFFQVE-QPSEAATVLN 243

Query: 241 NYQIEKKDNEYALTLKGSKMDNKADLLAGFYQDLVSLKISPSAIEVIDNSLEELYLEVT 299
            Y +  K N   + L   ++    +L       LV  +I    ++VI  SLE+ +LE+T
Sbjct: 244 QYDLLSKTNGVEIKLAKEEVPAVIEL-------LVMQQIRIYEVKVITKSLEDRFLEMT 295
```

There is also homology to SEQ ID 686.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 596

A DNA sequence (GBSx0636) was identified in *S.agalactiae* <SEQ ID 1847> which encodes the amino acid sequence <SEQ ID 1848>. Analysis of this protein sequence reveals the following:

```
Possible site: 34
>>> Seems to have no N-terminal signal sequence
```

```
----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4040(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAB71491 GB:U53767 ORF6 [Bacillus pumilus]
 Identities = 39/134 (29%), Positives = 71/134 (52%), Gaps = 16/134 (11%)

Query: 2   LGENIYLQRTQIGMTQENLSDYLHLTKTTISKWENNQAKPDIDYLILMANLFDISLDDLV 61
           LG NI  +R  + ++QE +++ L +++  ISKWE NQ++P +D LI +A LFD  + +LV
Sbjct: 4   LGSNISNKRKSLKLSQEYVAEQLGVSRQAISKWETNQSEPSMDNLIRLAELFDSDIKELV 63

Query: 62  GYQKTLSDDQRNQLIKDLKIKANVLSERDFFQEVKELSKQFPNDFKTLLIMINM--VLSN 119
                 S +Q ++  KDL+ +             K++  Q    F  +L++I+    +
Sbjct: 64  ------SPEQYSEEQKDLETRIE--------HGQKDIKMQMSAVFGRILMLISFFGYIGA 109

Query: 120 LTNLNDSEMKEWSL 133
           L +L+  ++  W L
Sbjct: 110 LFDLSSYQLPIWXL 123
```

There is also homology to SEQ ID 1740.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 597

A DNA sequence (GBSx0637) was identified in *S.agalactiae* <SEQ ID 1849> which encodes the amino acid sequence <SEQ ID 1850>. Analysis of this protein sequence reveals the following:

```
Possible site: 20
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-13.59   Transmembrane  152 - 168 ( 145 - 173)
   INTEGRAL    Likelihood = -9.71   Transmembrane    7 -  23 (   3 -  27)
   INTEGRAL    Likelihood = -6.95   Transmembrane  125 - 141 ( 122 - 146)
   INTEGRAL    Likelihood = -4.51   Transmembrane   85 - 101 (  83 - 102)
   INTEGRAL    Likelihood = -3.35   Transmembrane   55 -  71 (  54 -  75)

----- Final Results -----
               bacterial membrane --- Certainty=0.6434(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAA79986 GB:Z21972 ORF2 [Bacillus megaterium]
 Identities = 51/186 (27%), Positives = 106/186 (56%), Gaps = 5/186 (2%)

Query: 5   SFFQCVILLVSFLVLTLAVKSQSDMISYLDNITSAFFQSIRNPDLTNLMTIISTVVSPLT 64
           +F   V+ L+ F  +   + S ++ + + +++ S   Q   +P LT++M   + + S  +
Sbjct: 10  AFIISVLSLIGFSFMAFTI-SANEYLKFDEDVIS-LVQGWESPLLTDIMKFFTYIGSTAS 67

Query: 65  TSLIALVILGYQY-FLNQRIAVWLFM-LFFGTNALALLLKDIIARHRP-MNQLVFDSGYS 121
             +++LVIL + Y  L  R+ + LF  +  G+  L L++K    R RP +++L+   GYS
Sbjct: 68  LIILSLVILFFLYRILKHRLELVLFTAVMVGSPLLNLMVKLFFQRARPDLHRLIDIGGYS 127

Query: 122 FPSGHTISAFLLMILVLVVARQRLRRVLSQVVFVIFALVILASVIFSRLYLENHFLTDIL 181
           FPSGH ++AF L  ++  +  + +    ++++ ++F+++++ S+  SR+YL  H+ +DI+
Sbjct: 128 FPSGHAMNAFSLYGILTFLLWRHITARWARILLILFSMLMILSIGISRIYLGVHYPSDII 187

Query: 182 GSLLLG 187
              L G
Sbjct: 188 AGYLAG 193
```

There is also homology to SEQ ID 1852.

A related GBS gene <SEQ ID 8611> and protein <SEQ ID 8612> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 3
McG: Discrim Score:     11.91
GvH: Signal Score (-7.5): -4.6
     Possible site: 20
>>> Seems to have an uncleavable N-term signal seq
ALOM program   count: 5 value: -13.59 threshold:  0.0
   INTEGRAL    Likelihood =-13.59   Transmembrane  152 - 168 ( 145 - 173)
   INTEGRAL    Likelihood = -9.71   Transmembrane    7 -  23 (   3 -  27)
   INTEGRAL    Likelihood = -6.95   Transmembrane  125 - 141 ( 122 - 146)
   INTEGRAL    Likelihood = -4.51   Transmembrane   85 - 101 (  83 - 102)
   INTEGRAL    Likelihood = -3.35   Transmembrane   55 -  71 (  54 -  75)
   PERIPHERAL  Likelihood =  1.16     184
 modified ALOM score:   3.22

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.6434(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
ORF01359(313 - 864 of 1212)
EGAD|16772|16959(10 - 194 of 216) hypothetical protein {Bacillus megaterium}
GP|288301|emb|CAA79986.1||Z21972 ORF2 {Bacillus megaterium} PIR|S32217|S32217 hypothetical
protein 2 - Bacillus megaterium
%Match = 9.5
%Identity = 28.2  %Similarity = 60.1
Matches = 53  Mismatches = 68  Conservative Sub.s = 60

66        96       126       156       186       216       246       276
SFFIEFTHPFLIICNIHYSLRFKYIVAILLY**KFER*LIGKVRIWYFF*FVNSHI*T*KVSAYFKHFLNILNHNV*RFI

306       336       366       396       426       456       486       516
SLLK*GYVVNKKSFFQCVILLVSFLVLTLAVKSQSDMISYLDNITSAFFQSIRNPDLTNLMTIISTVVSPLTTSLIALVI
            :|   |: |: |  : : : | :: : : :::  :: |   :| ||::|   : : |  :  :::|||
   MKLKQQLTIAFIISVLSLIGFSFMAFTI-SANEYLKFDEDV-ISLVQGWESPLLTDIMKFFTYIGSTASLIILSLVI
           10        20         30        40         50        60        70

543       570       600       630       657       687       714       744
LGYQY-FLNQRIAVWLFM-LFFGTNALALLLKDIIARHRP-MNQLVFDSGYSFPSGHTISAFLLM-ILVLVVARQRLRRV
| : |  | :|: : ||  :  |:  | |::|    | || :::|:   |||||||| ::|| |  || ::: |: :
LFFLYRILKHRLELVLFTAVMVGSPLLNLMVKLFFQRARPDLHRLIDIGGYSFPSGHAMNAFSLYGILTFLLWRH-ITAR
             90       100       110       120       130       140       150

774       804       834       864       894       924       954       984
LSQVVFVIFALVILASVIFSRLYLENHFLTDILGSLLLGASSYYGLSAIVSLKELQ*K**LPMNYKRAFLKGSFIIHYFS
 :::::::|::::: |:  ||:||  |: :||:   | |
WARILLILFSMLMILSIGISRIYLGVHYPSDIIAGYLAGGCWIAISIWFFQRYQDRRKNKDR
             170       180       190       200       210
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 598

A DNA sequence (GBSx0638) was identified in *S.agalactiae* <SEQ ID 1853> which encodes the amino acid sequence <SEQ ID 1854>. Analysis of this protein sequence reveals the following:

```
Possible site: 41
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4288(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB15470 GB:Z99121 yvdC [Bacillus subtilis]
 Identities = 53/96 (55%), Positives = 70/96 (72%)

Query: 1  MDITDYQKWVSEFYKKRNWYQYNSFIRSNFLSEEVGELAQAIRKYEIGRDRPDETEQTDL 60
          M + D +KW+ EFY+KR W +Y  FIR  FL EE GELA+A+R YEIGRDRPDE E +
Sbjct: 1  MQLADAEKWMKEFYEKRGWTEYGPFIRVGFLMEEAGELARAVRAYEIGRDRPDEKESSRA 60

Query: 61 ENLNDIKEELGDVLDNIFILADQYNISLEEIISAHR 96
          E   ++ EE+GDV+ NI ILAD Y +SLE+++ AH+
Sbjct: 61 EQKQELIEEMGDVIGNIAILADMYGVSLEDVMKAHQ 96
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 599

A DNA sequence (GBSx0639) was identified in *S.agalactiae* <SEQ ID 1855> which encodes the amino acid sequence <SEQ ID 1856>. Analysis of this protein sequence reveals the following:

```
Possible site: 54
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0635(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAB06803 GB:AP001517 unknown conserved protein [Bacillus halodurans]
 Identities = 83/186 (44%), Positives = 117/186 (62%)

Query: 1   MRITIFCGASTGENPVYSEKTVALAQWMAQNKHSLVYGGGKVGLMGVMADTVIANGGYTT 60
           M+I +FCG+S G + VY E    L + +A+   +LVYGG  VG+MG +AD+V+  GG
Sbjct: 1   MKIAVFCGSSNGASDVYKEGARQLGKELARRGITLVYGGASVGIMGAVADSVLEAGGEVI 60

Query: 61  GVIPTFLRDREIAHENLSELIIVNNMPERKAKMMLLGDAFIALPGGPGTLEEISEVISWS 120
           GV+P FL + EI+H +L++LI+V  M ERKAKM  L D F+ALPGGPGTLEE  E+ +W+
Sbjct: 61  GVMPRFLEEPEISHPHLTKLIVVETMHERKAKMAELADGFLALPGGPGTLEEFFEIFTWA 120

Query: 121 RIGQNDNPCILYNVNGYFNDLKNMFDHMVGEGFLSLEDRENVLFSDDITEIEDFITNYKV 180
           +IG +  PC L N+N YF+ L  +  HM  E FL  + R   L   D   + D  + Y+
Sbjct: 121 QIGLHQKPCGLLNINHYFDPLVTLLHHMSNEQFLHEKYRSMALVHTDPILLLDQFSTYEP 180

Query: 181 PSTRQY 186
           P+ + Y
Sbjct: 181 PTVKAY 186
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 600

A DNA sequence (GBSx0640) was identified in *S.agalactiae* <SEQ ID 1857> which encodes the amino acid sequence <SEQ ID 1858>. Analysis of this protein sequence reveals the following:

```
Possible site: 19
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -7.86   Transmembrane  222 - 238 ( 214 - 239)
   INTEGRAL    Likelihood = -6.69   Transmembrane   39 -  55 (  36 -  58)
   INTEGRAL    Likelihood = -4.25   Transmembrane  266 - 282 ( 266 - 284)
   INTEGRAL    Likelihood = -1.28   Transmembrane  166 - 182 ( 166 - 182)
   INTEGRAL    Likelihood = -1.01   Transmembrane  190 - 206 ( 190 - 206)
   INTEGRAL    Likelihood = -0.96   Transmembrane   70 -  86 (  70 -  86)

----- Final Results -----
               bacterial membrane --- Certainty=0.4142(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB12420 GB:Z99107 ydiL [Bacillus subtilis]
 Identities = 40/132 (30%), Positives = 63/132 (47%), Gaps = 8/132 (6%)

Query: 107 ESQNYDATFNI-----LMISYSVVVGPFFEEVLYRGIVLNLL-SKYGKWFAIITSGILFG 160
           ES+N  A  ++     LMI  S +VGP  EE+++R I+   L  K   +FA + S ++FG
Sbjct: 114 ESENTQAILDVIQAVPLMIIVSSIVGPILEEIIFRKIIFGALYEKTNFFFAGLISSVIFG 173

Query: 161 LMHQDISQLLTTSIAGIIMGFI-AYHYSFKVALLLHICNNFIVEIFTQLSTVNELYGTYF 219
           ++H D+  LL  +  G    F+ A      V +  H+  N  V +  QL  V
Sbjct: 174 IVHADLKHLLLYTAMGFTFAFLYARTKRIWVPIFAHLMMNTFV-VIMQLEPVRNYLEQQS 232

Query: 220 ENILLILAILFI 231
             + LI+  LF+
Sbjct: 233 TQMQLIIGGLFL 244
```

No corresponding DNA sequence was identified in *S.pyogenes*.

A related GBS gene <SEQ ID 8613> and protein <SEQ ID 8614> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 6
McG: Discrim Score:     12.52
GvH: Signal Score (-7.5): -1.74
     Possible site: 19
>>> Seems to have a cleavable N-term signal seq.
ALOM program   count: 2 value:  -6.69 threshold:  0.0
   INTEGRAL    Likelihood = -6.69   Transmembrane   39 -  55 (  36 -  58)
   INTEGRAL    Likelihood = -0.96   Transmembrane   70 -  86 (  70 -  86)
   PERIPHERAL  Likelihood =  4.56      21
 modified ALOM score:   1.84

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.3675(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
Query:    10 LIGLILLAQAIVLSLATTLFAEILQNDVWIGIASTLIALLIPCF 53
             L+ L LL ++++LS++      +L   +W+ +A+ L+A ++ CF
Sbjct:    21 LLCLCLLVRSLLLSVSLYSALILLVLILWVTVATPLLAFVVSCF 64
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 601

A DNA sequence (GBSx0641) was identified in *S.agalactiae* <SEQ ID 1859> which encodes the amino acid sequence <SEQ ID 1860>. This protein is predicted to be capa protein. Analysis of this protein sequence reveals the following:

```
Possible site: 50
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-13.80    Transmembrane    27 -  43 (   22 -   50)

----- Final Results -----
              bacterial membrane --- Certainty=0.6519(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9385> which encodes amino acid sequence <SEQ ID 9386> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAF13661 GB:AF188935 pXO2-56 [Bacillus anthracis]
 Identities = 68/224 (30%), Positives = 118/224 (52%), Gaps = 10/224 (4%)

Query: 95  FKEVKSWIESADLAIGDYEGTISSE----YPLAGYPL-FNAPNEIATTMKETGYDVVDLA 149
           F+ V  +++++D   G++E  +  E    Y  A   +  +A  E    +KE G+ V++LA
Sbjct: 87  FRHVSPYLKNSDYVSGNFEHPVLLEDKKNYQKADKNIHLSAKEETVKAVKEAGFTVLNLA 146

Query: 150 HNHILDSQLAGAINTVKTFNRLGLDTIGVYLKDRNKEDILIKHVNGIKIAILGYSYGY-N 208
           +NH+ D    G  +T+K F    LD +G     ++ ++I+ ++VNG+++A LG++  +
Sbjct: 147 NNHMTDYGAKGTKDTIKAFKEADLDYVGAGENFKDVKNIVYQNVNGVRVATLGFTDAFVA 206

Query: 209 GMEANVSKSDYEKHMSDLDTKKIKQDIKKAEKEADITIVMPQMGIEYQKKPTTEQVMLYH 268
           G  A   +        D+  K+I +     +  AD+ +V    G EY  KP+  Q  L
Sbjct: 207 GAIATKEQPGSLSMNPDVLLKQISKAKDPKKGNADLVVVNTHWGEEYDNKPSPRQEALAK 266

Query: 269 SMIKWGADIIFGGHPHVVEPSEVIKKDGQKKFIIYSMGNFISNQ 312
           +M+  GADII G HPHV++  +V K+      I YS+GNF+ +Q
Sbjct: 267 AMVDAGADIIVGHHPHVLQSFDVYKQG----IIFYSLGNFVFDQ 306
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1861> which encodes the amino acid sequence <SEQ ID 1862>. Analysis of this protein sequence reveals the following:

```
Possible site: 45
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-12.05    Transmembrane    44 -  60 (   40 -   68)

----- Final Results -----
              bacterial membrane --- Certainty=0.5819(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related sequence was also identified in GAS <SEQ ID 9119> which encodes the amino acid sequence <SEQ ID 9120>. Analysis of this protein sequence reveals the following:

```
     Possible cleavage site: 31
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial membrane --- Certainty= 0.582(Affirmative) < succ>
               bacterial outside --- Certainty= 0.000(Not Clear) < succ>
```

```
                bacterial cytoplasm --- Certainty= 0.000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
  Identities = 232/334 (69%), Positives = 273/334 (81%), Gaps = 4/334 (1%)

Query: 24  YQKTLIFCVAVIIAIFILGLSKDLAQSKGQKVANNNT----VKTARVVANGDILLHDVLY 79
            Y+KT+   VA+I+A+ + GL  DL   +  ++A   +    VKTARVVANGDIL+HD+LY
Sbjct: 40  YKKTMATVVALIVALLLFGLIYDLLGVQKNELAAQKSAQPKVKTARVVANGDILIHDILY 99

Query: 80  ASARQPDGTYNFTPYFKEVKSWIESADLAIGDYEGTISSEYPLAGYPLFNAPNEIATTMK 139
             SAR+ D TY+FTPYF+ VK WI  ADLAIGDYEGTIS +YPLAGYPLFNAP EIA  +K
Sbjct: 100 MSARKADDTYDFTPYFEYVKDWISGADLAIGDYEGTISPDYPLAGYPLFNAPEEIAGALK 159

Query: 140 ETGYDVVDLAHNHILDSQLAGAINTVKTFNRLGLDTIGVYLKDRNKEDILIKHVNGIKIA 199
             TGYDVVDLAHNHILDSQL GA+NT K F++LG+D+IG+Y KDR+KE  LIK+VNGIKIA
Sbjct: 160 NTGYDVVDLAHNHILDSQLDGALNTKKVFHQLGIDSIGIYDKDRSKESFLIKNVNGIKIA 219

Query: 200 ILGYSYGYNGMEANVSKSDYEKHMSDLDTKKIKQDIKKAEKEADITIVMPQMGIEYQKKP 259
            ILGYSYGYNGMEA +S+ DYEKHMSDLD  KIK++++ AEK+AD+TIVMPQMG EY  +P
Sbjct: 220 ILGYSYGYNGMEATLSQEDYEKHMSDLDEAKIKKELQLAEKKADVTIVMPQMGTEYALEP 279

Query: 260 TTEQVMLYHSMIKWGADIIFGGHPHVVEPSEVIKKDGQKKFIIYSMGNFISNQRLETVDD 319
            T EQ  LYH MI WGAD++ GGHPHV+EPSE + K  QKKFIIYSMGNFISNQRLETVDD
Sbjct: 280 TAEQKELYHKMIDWGADVVLGGHPHVIEPSETVIKGRQKKFIIYSMGNFISNQRLETVDD 339

Query: 320 IWTERGLLMDVTIEKKGQKTVIKKVKAHPTLVEA 353
            IWTERGLLMD+T EKK  KT IK V+AHPT+V A
Sbjct: 340 IWTERGLLMDLTFEKKDNKTKIKTVEAHPTMVLA 373
```

A related GBS gene <SEQ ID 8615> and protein <SEQ ID 8616> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop Possible site: -1   Crend: 7
SRCFLG: 0
McG: Length  of  UR:    18
     Peak Value of UR:   3.83
     Net Charge of CR: 2
McG: Discrim Score:     15.36
GvH: Signal Score (-7.5): -1.52
     Possible site: 32
>>> Seems to have a cleavable N-term signal seq.
Amino Acid Composition: calculated from 33
ALOM program   count: 0 value:   4.35 threshold:  0.0
   PERIPHERAL  Likelihood =  4.35     170
 modified ALOM score:  -1.37

*** Reasoning Step: 3

Rule gpol

----- Final Results -----
                bacterial outside --- Certainty=0.3000(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
30.6/53.3% over 230aa
                                                              Bacillus anthracis
  EGAD|20151| capa protein Insert characterized
  SP|P19579|CAPA_BACAN CAPA PROTEIN. Edit characterized
  GP|142633|gb|AAA22288.1||M24150 46 Kd encapsulation protein CapA Insert characterized
  PIR|C30091|C30091 capA protein - Insert characterized

ORF02075(574 - 1257 of 1734)
```

```
EGAD|20151|20674(83 - 313 of 411) capa protein {Bacillus anthracis} SP|P19579|CAPA_BACAN
CAPA PROTEIN. GP|142633|gb|AAA22288.1||M24150 46 Kd encapsulation protein CapA {Bacillus
anthracis} PIR|C30091|C30091 capA protein - Bacillus anthracis
%Match = 8.9
%Identity = 30.6  %Similarity = 53.3
Matches = 70  Mismatches = 102  Conservative Sub.s = 52

468       498       528       558       585       615       645       663
LAQSKGQKVANNNTVKTARVVANGDILLHDVLYASARQPDGTYNFTPY-FKEVKSWIESADLAIGDYEGTI----SSEYP
:| :  |:      ||          :: |::     :       | | |: |  :::::|   |::|  :       |
IAATWVQRTEAVAPVKHRENEKLTMTMVGDIMMGRHVKEIVNRYGTDYVFRHVSPYLKNSDYVSGNFEHPVLLEDKKNYQ
           50        60        70        80        90       100       110

690       720       750       780       810       840       870       900
LAGYPL-FNAPNEIATTMKETGYDVVDLAHNHILDSQLAGAINTVKTFNRLGLDTIGVYLKDRNKEDILIKHVNGIKIAI
 |   : ::|  |    :|| |: |::||:||: |    |  :|:| |    || :|     :: ::|: ::|||:::|
KADKNIHLSAKEETVKAVKEAGFTVLNLANNHMTDYGAKGTKDTIKAFKEADLDYVGAGENFKDVKNIVYQNVNGVRVAT
          130       140       150       160       170       180       190

927       957       987      1017      1047      1077      1107      1137
LGYSYGY-NGMEANVSKSDYEKHMSDLDTKKIKQDIKKAEKEADITIVMPQMGIEYQKKPTTEQVMLYHSMIKWGADIIF
||::  :  |  |   :        |:  |:| :     :  ||: :|  : | ||  ||:  |  |  :|:  |||||
LGFTDAFVAGAIATKEQPGSLSMNPDVLLKQISKAKDPKKGNADLVVVNTHWGEEYDNKPSPRQEALAKAMVDAGADIIV
          210       220       230       240       250       260       270

1167      1197      1227      1257      1287      1317      1347      1377
GGHPHVVEPSEVIKKDGQKKFIIYSMGNFISNQRLETVDDIWTERGLLMDVTIEKKGQKTVIKKVKAHPTLVEAKPNGRY
| ||||::  :| |    :  | ||:|||: :|      |    :  | |          :     :   :  |    |
GHHPHVLQSFDVYK----QGIIFYSLGNFVFDQGWTRTKDSALVQYHLRDNGTAILDVVPLNIQEGSPKPVASALDKNRV
          290           300       310       320       330       340       350
```

SEQ ID 8616 (GBS289) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 57 (lane 5; MW 40kDa), in Figure 181 (lane 6; MW 47kDa), in Figure 169 (lane 13 & 14; MW 54.5kDa – thioredoxin fusion) and in Figure 239 (lane 3; MW 54.5kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 61 (lane 5; MW 65kDa).

SEQ ID 8616 (GBS289L) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 126 (lane 2; MW 72kDa) and in Figure 184 (lane 5; MW 72kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 126 (lane 5-7; MW 47kDa).

GBS289L-His was purified as shown in Figure 234, lane 9-10. Purified GBS289L-GST is shown in Figure 245, lane 10.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 602

A DNA sequence (GBSx0642) was identified in *S.agalactiae* <SEQ ID 1863> which encodes the amino acid sequence <SEQ ID 1864>. This protein is predicted to be thiamin biosynthesis protein ThiI (thiI). Analysis of this protein sequence reveals the following:

```
Possible site: 55
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
                bacterial cytoplasm --- Certainty=0.2720(Affirmative) < succ>
                 bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
```

```
                  bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9971> which encodes amino acid sequence <SEQ ID 9972> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAC00308 GB:AF008220 YtbJ [Bacillus subtilis]
 Identities = 184/354 (51%), Positives = 249/354 (69%)

Query: 11  MQYSEIMIRYGELSTKKKNRMRFINKLKNNMEHVLSIYPDVSVKTDRDRGHVYLNGTDYH 70
            M Y  I+IR+GE+STK KNR  FI +LK N+  VL  YP++   ++RDR  + LNG D
Sbjct: 1   MNYDHILIRFGEISTKGKNRKSFIERLKQNIRLVLKDYPNLKYFSNRDRMTITLNGEDPE 60

Query: 71  EVAESLKEIFGIQAFSPSFKVEKNVDTLVKAVQEIMTSVYKDGMTFKITAKRSDHSFELD 130
             +   LK++FGIQ+FS + K +  +D +     + +   YK G TFK+  KR+   FELD
Sbjct: 61  ALFPHLKQVFGIQSFSLAIKCDSRLDDIKATALKAIKDQYKPGDTFKVATKRAYKQFELD 120

Query: 131 SRALNHTLGDAVFSVLPNIKAQMKQPDINLKVEIRDEAAYISYEDIRGAGGLPVGTSGKG 190
            +  +N  +G  +      +   ++ PDI L++EIR+EA +++  D +GAGGLPVG++GK
Sbjct: 121 TNQMNAEIGGHILRNTEGLTVDVRNPDIPLRIEIREEATFLTIRDEKGAGGLPVGSAGKA 180

Query: 191 MLMLSGGIDSPVAGYLALKRGVDIEAVHFASPPYTSPGALKKAHDLTRKLTKFGGNIQFI 250
            MLMLSGG DSPVAG+ A+KRG+ +EAVHF SPPYTS  A +K  DL + L++FGG++
Sbjct: 181 MLMLSGGFDSPVAGFYAMKRGLSVEAVHFFSPPYTSERAKQKVMDLAKCLSRFGGSMTLH 240

Query: 251 EVPFTEIQEEIKAKAPEAYLMTLTRRFMMRITDRIREDRNGLVIINGESLGQVASQTLES 310
             VPFT+ QE I+ + PE Y MT TRR M++I DRIRE RNGL II GESLGQVASQTLES
Sbjct: 241 IVPFTKTQELIQKQIPENYTMTATRRLMLQIADRIREKRNGLAIITGESLGQVASQTLES 300

Query: 311 MQAINAVTATPIIRPVVTMDKLEIIDIAQKIDTFDISIQPFEDCCTIFAPDRPK 364
            M AINAVT+TPI+RP++ MDK EII+ +++I T++ SIQPFEDCCTIF   +P+
Sbjct: 301 MYAINAVTSTPILRPLIAMDKTEIIEKSREIGTYETSIQPFEDCCTIFTTAKPR 354
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1865> which encodes the amino acid sequence <SEQ ID 1866>. Analysis of this protein sequence reveals the following:

```
Possible site: 42
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4897(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 316/404 (78%), Positives = 362/404 (89%)

Query: 11  MQYSEIMIRYGELSTKKKNRMRFINKLKNNMEHVLSIYPDVSVKTDRDRGHVYLNGTDYH 70
            M YSEIM+R+GELSTK KNRMRFINKLKNN++ VL+ +P ++V++DRDR HV LNGTDY
Sbjct: 1   MDYSEIMVRHGELSTKGKNRMRFINKLKNNIQDVLAPFPAITVRSDRDRTHVSLNGTDYQ 60

Query: 71  EVAESLKEIFGIQAFSPSFKVEKNVDTLVKAVQEIMTSVYKDGMTFKITAKRSDHSFELD 130
             + E+LK +FG+QA SP +K+EK+V  LV AVQ+IMTS+Y+DG+TFKI  KRSDH+FELD
Sbjct: 61  PIVEALKLVFGVQALSPVYKLEKSVPLLVTAVQDIMTSLYRDGLTFKIATKRSDHAFELD 120

Query: 131 SRALNHTLGDAVFSVLPNIKAQMKQPDINLKVEIRDEAAYISYEDIRGAGGLPVGTSGKG 190
            SR LN  LG AVF VLPNI+AQMK PD+ LKVEIRDEAAYISYE+I+GAGGLPVGTSGKG
Sbjct: 121 SRELNSLLGGAVFEVLPNIQAQMKHPDVTLKVEIRDEAAYISYEEIKGAGGLPVGTSGKG 180

Query: 191 MLMLSGGIDSPVAGYLALKRGVDIEAVHFASPPYTSPGALKKAHDLTRKLTKFGGNIQFI 250
            MLMLSGGIDSPVAGYLALKRG+DIE VHFASPPYTSPGAL KA DLTR+LT+FGGNIQFI
Sbjct: 181 MLMLSGGIDSPVAGYLALKRGLDIEVVHFASPPYTSPGALAKAQDLTRRLTRFGGNIQFI 240

Query: 251 EVPFTEIQEEIKAKAPEAYLMTLTRRFMMRITDRIREDRNGLVIINGESLGQVASQTLES 310
```

```
              EVPFTEIQEEIK KAPEAYLMTLTRRFMMRITD IRE R GLVI+NGESLGQVASQTLES
Sbjct: 241 EVPFTEIQEEIKNKAPEAYLMTLTRRFMMRITDAIREQRKGLVIVNGESLGQVASQTLES 300

Query: 311 MQAINAVTATPIIRPVVTMDKLEIIDIAQKIDTFDISIQPFEDCCTIFAPDRPKTNPKIK 370
           MQAINAVT+TPIIRPVVTMDKLEII++AQ IDTFDISIQPFEDCCTIFAPDRPKTNPK+
Sbjct: 301 MQAINAVTSTPIIRPVVTMDKLEIIEMAQAIDTFDISIQPFEDCCTIFAPDRPKTNPKLG 360

Query: 371 NTEQYEKRMDVEGLVERAVAGIMVTTIQPQADSDDVDDLIDDLL 414
           N E+YE+  D++GLV+RAV+GI+VT I P+  +D+V++LID LL
Sbjct: 361 NAEKYEECFDIDGLVQRAVSGIVVTEITPEIVNDEVENLIDALL 404
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 603

A DNA sequence (GBSx0643) was identified in *S.agalactiae* <SEQ ID 1867> which encodes the amino acid sequence <SEQ ID 1868>. This protein is predicted to be nifs protein homolog , fragment. Analysis of this protein sequence reveals the following:

```
Possible site: 47
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.27   Transmembrane  131 - 147 ( 131 - 147)

----- Final Results -----
               bacterial membrane --- Certainty=0.1107(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAA43493 GB:X61190 nifS-like gene [Lactobacillus delbrueckii]
 Identities = 177/353 (50%), Positives = 234/353 (66%), Gaps = 1/353 (0%)

Query: 14  PEVLRTYQEVASKIYGNPSSLHELGTTSSRILEASRKQIASLLELKANEIFFTSGGTEAD 73
            P+ L TY +V +KI+GNPSSLH+LG  +  +LEASRKQ+A LL +  +EI+FTSGGTE++
Sbjct: 3   PKALETYSQVVTKIWGNPSSLHKLGDRAHGLLEASRKQVADLLGVNTDEIYFTSGGTESN 62

Query: 74  NWVIKGLAFEKQHFGNHIIVSDIEHPAVKESAKWLGEYGFEIDYAPVDDKGFVDVEALVK 133
            N  IKG A+ K+ FG HII S +EH +V  +   L   GF +   PVD +G V+ E L
Sbjct: 63  NTAIKGTAWAKREFGKHIITSSVEHASVANTFTELENLGFRVTRLPVDKEGRVNPEDLKA 122

Query: 134 LIKPETILISIMAINNEIGSIQPIKAISDLLSDKPTISFHVDAVQAIGKIPTKDYLTERV 193
             +  +T L+SIM +NNEIG+IQPIK IS++L+D P I FHVD VQA+GK       T RV
Sbjct: 123 ALDKDTTLVSIMGVNNEIGTIQPIKEISEILADYPNIHFHVDNVQALGKGIWDQVFTSRV 182

Query: 194 DFASFSSHKFHGVRGVGFLYIKEGKRISPLLTGGGQETDLRSTTENVAGIAATAKALRMV 253
            D  SFSSHKFHG RG+G LY K G+ + PL  GGGQE  LRS TEN+A IAA AKA R++
Sbjct: 183 DMMSFSSHKFHGPRGIGILYKKRGRMLMPLCEGGGQEKGLRSGTENLAAIAAMAKAARLL 242

Query: 254 MDKEVVAIPKISKMKTIIHDELAKYEDITLFSG-KEDFSPNIITFGIKGVRGEVLVHAFE 312
            +  E      +   +K  I    LA     I +FS  K DF+P+I+ F ++G+RGE LVH  E
Sbjct: 243 LTDEKEKADREYAIKEKISKYLAGKPGIHIFSPLKADFAPHILCFALEGIRGETLVHTLE 302

Query: 313 GHDIFISTTSACSSKAGKPAGTLIAMGISTKLAQTAVRISLDDDNDMGQVEQF 365
              DI+ISTTSAC+SK    A TL+AM      +A +AVR+S D+ N + + ++F
Sbjct: 303 DQDIYISTTSACASKKADEASTLVAMKTPDAIATSAVRLSFDESNTLEEADEF 355
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1869> which encodes the amino acid sequence <SEQ ID 1870>. Analysis of this protein sequence reveals the following:

```
Possible site: 41
>>> Seems to have no N-terminal signal sequence
```

```
----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3067(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 268/370 (72%), Positives = 322/370 (86%)

Query: 1   MIYFDNSATTIPYPEVLRTYQEVASKIYGNPSSLHELGTTSSRILEASRKQIASLLELKA 60
           MIYFDN+ATTIPY E L+TYQEVA+KIYGNPSSLH+LGT +SRILEASRKQIA LL +K+
Sbjct: 1   MIYFDNAATTIPYGEALKTYQEVATKIYGNPSSLHQLGTNASRILEASRKQIAGLLGVKS 60

Query: 61  NEIFFTSGGTEADNWVIKGLAFEKQHFGNHIIVSDIEHPAVKESAKWLGEYGFEIDYAPV 120
            EIFFTSGGTE+ NW IKG+AFEK  FG HII+S IEHPAV ES KWL   GFE+ YAPV
Sbjct: 61  EEIFFTSGGTESANWAIKGIAFEKNAFGKHIIISAIEHPAVSESVKWLLTQGFEVSYAPV 120

Query: 121 DDKGFVDVEALVKLIKPETILISIMAINNEIGSIQPIKAISDLLSDKPTISFHVDAVQAI 180
             +G VDV AL +LI+P+TILISIMA+NNE+G+IQPI+AIS+LL+++PTI+FHVDAVQAI
Sbjct: 121 TTQGVVDVNALAELIRPDTILISIMAVNNEMGAIQPIRAISNLLANQPTITFHVDAVQAI 180

Query: 181 GKIPTKDYLTERVDFASFSSHKFHGVRGVGFLYIKEGKRISPLLTGGGQETDLRSTTENV 240
           GKIP  DY+T RVD ASFS HKFH VRGVGFLY K GKR++PLL+GGGQE +LRSTTENV
Sbjct: 181 GKIPLCDYMTNRVDLASFSGHKFHSVRGVGFLYKKAGKRLNPLLSGGGQEQELRSTTENV 240

Query: 241 AGIAATAKALRMVMDKEVVAIPKISKMKTIIHDELAKYEDITLFSGKEDFSPNIITFGIK 300
           AGIA+ AKALR+V +K+V  +PK++ M+ +I+  L+ Y D+T+FS +E F+PNI+TFGI+
Sbjct: 241 AGIASMAKALRIVTEKQVSVLPKLTAMRDVIYKSLSAYPDVTVFSAQEGFAPNILTFGIR 300

Query: 301 GVRGEVLVHAFEGHDIFISTTSACSSKAGKPAGTLIAMGISTKLAQTAVRISLDDDNDMG 360
           GVRGEV+VHAFE ++I+ISTTSACSSKAG+PAG+L+AMGI   K AQTAVRISLDDDNDMG
Sbjct: 301 GVRGEVIVHAFEKYEIYISTTSACSSKAGEPAGSLVAMGIPVKTAQTAVRISLDDDNDMG 360

Query: 361 QVEQFLTIFK 370
           QVEQFLTIF+
Sbjct: 361 QVEQFLTIFQ 370
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 604

A DNA sequence (GBSx0644) was identified in *S.agalactiae* <SEQ ID 1871> which encodes the amino acid sequence <SEQ ID 1872>. Analysis of this protein sequence reveals the following:

```
Possible site: 29
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1539(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 605

A DNA sequence (GBSx0645) was identified in *S.agalactiae* <SEQ ID 1873> which encodes the amino acid sequence <SEQ ID 1874>. This protein is predicted to be glutathione reductase (gor). Analysis of this protein sequence reveals the following:

```
Possible site: 23
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -4.25    Transmembrane  170 - 186 ( 169 - 187)

----- Final Results -----
               bacterial membrane --- Certainty=0.2699(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAA76640 GB:AB019579 glutathione reductase (GR) [Streptococcus mutans]
 Identities = 274/450 (60%), Positives = 346/450 (76%), Gaps = 1/450 (0%)

Query: 1   MSKQYDYIVIGGGSAGSGTANRAAMYGAKVLLIEGGQVGGTCVNLGCVPKKIMWYGAQVS 60
           M+KQYDYIVIGGGS G  +ANRAAM+GAKV+L EG QVGGTCVN+GCVPKK+MWYGAQV+
Sbjct: 1   MTKQYDYIVIGGGSGGIASANRAAMHGAKVILFEGKQVGGTCVNVGCVPKKVMWYGAQVA 60

Query: 61  ETLHKYSSGYGFEVNNLNFDFTTLKANRDAYVQRSRQSYAANFERNGVEKIDGFARFIDN 120
           ET++ Y++ YGF+V    F F  LK NR AY+ R + SY   F+ NGVE++   +A F+D
Sbjct: 61  ETINNYAADYGFDVTTQTFHFDALKQNRQAYIDRIQDSYERGFDSNGVERVYSYATFVDA 120

Query: 121 HTIEVNGQQYKAPHITIATGGHPLYPDIIGSELGETSDDFFGWETLPDSILIVGAGYIAA 180
           HT+EV G+ Y APHI IATGGH L PDI GSE G TSD FF  + +P    +VGAGYIA
Sbjct: 121 HTVEVAGEHYTAPHILIATGGHALLPDIPGSEYGITSDGFFELDAIPKRTAVVGAGYIAV 180

Query: 181 ELAGVVNELGVETHLAFRKDHILRGFDDMVTSEVMAEMEKSGISLHANHVPKSLKRDEGG 240
           E++GV++ LG ETHL  R+D  LR FD  +   ++ EM+K G  LH   VPK + ++
Sbjct: 181 EISGVLHALGGETHLFVRRDRPLRKFDKEIVGTLVDEMKKDGPHLHTFSVPKEVIKNTDN 240

Query: 241 KLIFEAENGKTLVVDRVIWAIGRGPNV-DMGLENTDIVLNDKGYIKADEFENTSVDGVYA 299
            L    ENG+   VD +IWAIGR  N     LE T + L+ +G+I  D FENT+V+G+YA
Sbjct: 241 SLTLILENGEEYTVDTLIWAIGRAANTKGFNLEVTGVTLDSRGFIATDAFENTNVEGLYA 300

Query: 300 IGDVNGKIALTPVAIAAGRRLSERLFNHKDNEKLDYHNVPSVIFTHPVIGTVGLSEAAAI 359
           +GDVNGK+ LTPVA+ AGR+LSERLFNHK   K+DY +V +VIF+HPVIG++GLSE  A+
Sbjct: 301 LGDVNGKLELTPVAVKAGRQLSERLFNHKPQAKMDYKDVATVIFSHPVIGSIGLSEEVAL 360

Query: 360 EQFGEDNIKVYTSTFTSMYTAVTTNRQAVKMKLITLGKEEKVIGLHGVGYGIDEMIQGFS 419
           +Q+GE+N+ VY STFTSMYTAVT++RQA KMKL+T+G++EK++GLHG+GYG+DEMIQGF+
Sbjct: 361 DQYGEENVTVYRSTFTSMYTAVTSHRQACKMKLVTVGEDEKIVGLHGIGYGVDEMIQGFA 420

Query: 420 VAIKMGATKADFDDTVAIHPTGSEEFVTMR 449
           VAIKMGATKADFD+TVAIHPTGSEEFVTMR
Sbjct: 421 VAIKMGATKADFDNTVAIHPTGSEEFVTMR 450
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1875> which encodes the amino acid sequence <SEQ ID 1876>. Analysis of this protein sequence reveals the following:

```
Possible site: 23
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -1.33    Transmembrane  173 - 189 ( 173 - 191)

----- Final Results -----
               bacterial membrane --- Certainty=0.1532(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 268/446 (60%), Positives = 340/446 (76%), Gaps = 1/446 (0%)

Query: 5   YDYIVIGGGSAGSGTANRAAMYGAKVLLIEGGQVGGTCVNLGCVPKKIMWYGAQVSETLH 64
           YDYIVIGGGSAG  +ANRAAM+GAKVLL EG ++GGTCVNLGCVPKK+MWYGAQV++ L
Sbjct: 8   YDYIVIGGGSAGIASANRAAMHGAKVLLAEGKEIGGTCVNLGCVPKKVMWYGAQVADILG 67

Query: 65  KYSSGYGFEVNNLNFDFTTLKANRDAYVQRSRQSYAANFERNGVEKIDGFARFIDNHTIE 124
            Y+  YGF+     FDF  LKANR AY+ R   SY   FE+NGV++I  +A F D HT+E
Sbjct: 68  TYAKDYGFDFKEKAFDFKQLKANRQAYIDRIHASYERGFEQNGVDRIYDYAVFKDAHTVE 127

Query: 125 VNGQQYKAPHITIATGGHPLYPDIIGSELGETSDDFFGWETLPDSILIVGAGYIAAELAG 184
           + GQ Y APHI IATGGHP++PDI G++ G +SD FF  + +P    +VGAGYIA ELAG
Sbjct: 128 IAGQLYTAPHILIATGGHPVFPDIEGAQYGISSDGFFALDEVPKRTAVVGAGYIAVELAG 187

Query: 185 VVNELGVETHLAFRKDHILRGFDDMVTSEVMAEMEKSGISLHANHVPKSLKRDEGGKLIF 244
           V++ LG +T L  R D  LR FD  +   ++ EM  +G  LH +     + ++    L
Sbjct: 188 VLHALGSKTDLFIRHDRPLRSFDKTIVDVLVDEMAVNGPRLHTHAEVAKVVKNTDESLTL 247

Query: 245 EAENGKTLVVDRVIWAIGRGPNVD-MGLENTDIVLNDKGYIKADEFENTSVDGVYAIGDV 303
             ++G+ + VD++IWAIGR PN++   L+ T + LNDKGYI+ D +ENTSV G+YA+GDV
Sbjct: 248 YLKDGQEVEVDQLIWAIGRKPNLEGFSLDKTGVTLNDKGYIETDAYENTSVKGIYAVGDV 307

Query: 304 NGKIALTPVAIAAGRRLSERLFNHKDNEKLDYHNVPSVIFTHPVIGTVGLSEAAAIEQFG 363
           NGK+ALTPVA+AAGRRLSERLFN K +EKLDY NV +VIF+HPVIG+VGLSE AA++Q+G
Sbjct: 308 NGKLALTPVAVAAGRRLSERLFNGKTDEKLDYQNVATVIFSHPVIGSVGLSEEAAVKQYG 367

Query: 364 EDNIKVYTSTFTSMYTAVTTNRQAVKMKLITLGKEEKVIGLHGVGYGIDEMIQGFSVAIK 423
           ++ +K Y S FTSM+TA+T +RQ   MKL+T+G  EK++GLHG+GYG+DEMIQGF+VAIK
Sbjct: 368 QEAVKTYQSRFTSMFTAITNHRQPCLMKLVTVGDTEKIVGLHGIGYGVDEMIQGFAVAIK 427

Query: 424 MGATKADFDDTVAIHPTGSEEFVTMR 449
           MGATKADFD+TVAIHPTGSEEFVTMR
Sbjct: 428 MGATKADFDNTVAIHPTGSEEFVTMR 453
```

SEQ ID 1874 (GBS417) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 79 (lane 5; MW 53kDa).

GBS417-His was purified as shown in Figure 216, lane 2.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 606

A DNA sequence (GBSx0646) was identified in *S.agalactiae* <SEQ ID 1877> which encodes the amino acid sequence <SEQ ID 1878>. Analysis of this protein sequence reveals the following:

```
Possible site: 35
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3122(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAC62417 GB:AF084104 hypothetical protein [Bacillus firmus]
 Identities = 33/110 (30%), Positives = 66/110 (60%)

Query: 1   MANVYDLANELERAVRALPEYQAVLTAKSAIESDADAQVLWQDFLATQSKVQEMMQSGQM 60
           M+NVYD A+EL++A+    E+ A+ +    IE+D  A+ + ++F   Q ++Q+    G
Sbjct: 1   MSNVYDKAHELKKAIAESEEFSALKSMHEEIEADEIAKKMLENFRNLQLELQQKQMQGIQ 60

Query: 61  PSQEEQDEMSKLGEKIESNDLLKVYFDQQQRLSVYMSDIEKIVFAPMQDL 110
```

```
            ++EE  +  +  E ++ ++L+    + +QRLSV + DI KI+  P++++
Sbjct: 61  ITEEEAQKAQQQFELVQQHELISKLMEAEQRLSVIIGDINKIITEPLEEI 110
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1879> which encodes the amino acid sequence <SEQ ID 1880>. Analysis of this protein sequence reveals the following:

```
Possible site: 38
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4058(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 68/108 (62%), Positives = 86/108 (78%)

Query: 4   VYDLANELERAVRALPEYQAVLTAKSAIESDADAQVLWQDFLATQSKVQEMMQSGQMPSQ 63
           +YD AN+LERAVRALPEYQ VL  K AI++D  A  L+ +F+A Q K+Q MMQSGQMP+
Sbjct: 5   IYDYANQLERAVRALPEYQKVLEVKEAIQADVSASELFDEFVAMQEKIQGMMQSGQMPTA 64

Query: 64  EEQDEMSKLGEKIESNDLLKVYFDQQQRLSVYMSDIEKIVFAPMQDLM 111
           EEQ  + +L +KIE+ND LK YF+ QQ LSVYMSDIE+IVFAP++DL+
Sbjct: 65  EEQTSIQELSQKIEANDQLKAYFEAQQALSVYMSDIERIVFAPLKDLV 112
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 607

A DNA sequence (GBSx0647) was identified in *S.agalactiae* <SEQ ID 1881> which encodes the amino acid sequence <SEQ ID 1882>. This protein is predicted to be chorismate synthase (aroC). Analysis of this protein sequence reveals the following:

```
Possible site: 15
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -4.67   Transmembrane  343 - 359 ( 341 - 364)

----- Final Results -----
               bacterial membrane --- Certainty=0.2869(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAB05375 GB:AP001512 chorismate synthase [Bacillus halodurans]
 Identities = 227/381 (59%), Positives = 282/381 (73%), Gaps = 2/381 (0%)

Query: 1   MRYLTAGESHGPSLTAIIEGIPAGLKLSAKDINEDLKRRQGGYGRGNRMKIETDQVIISS 60
           MRYLTAGESHGP LT IIEG PA L+L A DIN DL RRQGG+GRG RM+IE DQV I
Sbjct: 1   MRYLTAGESHGPQLTTIIEGAPAQLELVADDINVDLARRQGGHGRGRRMQIEKDQVQIVG 60

Query: 61  GVRHGKTLGSPITLTVTNKDHSKWLDIMSVEDI--EERLKQKRRIKHPRPGHADLVGGIK 118
           G+RHGKT G+PI L V NKD   W  IM  E +  +E  + KR+I  PRPGHADL G IK
Sbjct: 61  GIRHGKTTGAPIALVVENKDWKHWTKIMGAEPLTGDEEKEIKRKITRPRPGHADLNGAIK 120

Query: 119 YRFDDLRNALERSSARETTMRVAIGAIAKRILKEIGIEIANHIVVFGGKEITVPDKLTVQ 178
           Y   D+RN LERSSARETT+RVA GA+AK+IL+  GIE+ +H++  GG +        +
Sbjct: 121 YGHRDMRNVLERSSARETTVRVAAGAVAKKILRTFGIEVGSHVLEIGGVKAEKTSYDQLS 180

Query: 179 QIKVLSSQSQVAIVNPSFEQEIKDYIDSVKKAGDTIGGVVETIVGGVPVGLGSYVHWDRK 238
            +K L+  S V  ++   EQE+   ID  K+ GD+IGGVVE IV GVP+GLGS+VH+DRK
Sbjct: 181 NLKELAEASPVRCLDKEAEQEMIAAIDQAKENGDSIGGVVEVIVEGVPIGLGSHVHYDRK 240
```

```
Query: 239 LDAKIAQAVVSINAFKGVEFGLGFKSGFLKGSQVMDSISWTKDQGYIRQSNNLGGFEGGM 298
           LDAKIA AV+SINAFKGVEFG+GF++    GS+V D I+W +++GY R+SNNLGGFEGGM
Sbjct: 241 LDAKIAAAVMSINAFKGVEFGIGFEAASKPGSEVHDEIAWDEERGYYRKSNNLGGFEGGM 300

Query: 299 TNGEPIIVRGVMKPIPTLYKPLMSVDIDTHEPYRATVERSDPTALPAAGVVMEAVVATVL 358
           TNG PI+VRGVMKPIPTLYKPL SVDI T EP+ A++ERSD  A+PAA VV EAVVA  +
Sbjct: 301 TNGMPIVVRGVMKPIPTLYKPLQSVDIATKEPFAASIERSDSCAVPAAAVVAEAVVAWEV 360

Query: 359 VTEVLEKFSSDNMYELKEAVK 379
              +LE+F +D + E+++ ++
Sbjct: 361 ANALLERFGADQVEEIEKNIR 381
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1883> which encodes the amino acid sequence <SEQ ID 1884>. Analysis of this protein sequence reveals the following:

```
     Possible site: 15
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.75   Transmembrane  342 - 358 ( 342 - 359)
   INTEGRAL    Likelihood = -0.16   Transmembrane  155 - 171 ( 155 - 171)

----- Final Results -----
               bacterial membrane --- Certainty=0.1298(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:BAB05375 GB:AP001512 chorismate synthase [Bacillus halodurans]
 Identities = 213/390 (54%), Positives = 277/390 (70%), Gaps = 2/390 (0%)

Query: 1   LRYLTAGESHGPSLTAIIEGIPAGLTLHPADIDHELQRRQGGYGRGARMSIETDRVQISS 60
           +RYLTAGESHGP LT IIEG PA L L   DI+ +L RRQGG+GRG RM IE D+VQI
Sbjct: 1   MRYLTAGESHGPQLTTIIEGAPAQLELVADDINVDLARRQGGHGRGRRMQIEKDQVQIVG 60

Query: 61  GVRHGKTTGAPITLTVINKDHQKWLDVMAVGDI--EETLKLKRRVKHPRPGHADLVGGIK 118
           G+RHGKTTGAPI L V NKD + W  +M    +  +E  ++KR++  PRPGHADL G IK
Sbjct: 61  GIRHGKTTGAPIALVVENKDWKHWTKIMGAEPLTGDEEKEIKRKITRPRPGHADLNGAIK 120

Query: 119 YHFNDLRDALERSSARETTMRVAVGAVAKRILAELGIDMLHHILIFGGITITIPSKLSFR 178
           Y   D+R+ LERSSARETT+RVA GAVAK+IL   GI++  H+L  GG+     S
Sbjct: 121 YGHRDMRNVLERSSARETTVRVAAGAVAKKILRTFGIEVGSHVLEIGGVKAEKTSYDQLS 180

Query: 179 ELQERALHSELSIVNPKQEEEIKTYIDKIKKEGDTIGGIIETIVQGVPAGLGSYVQWDKK 238
            L+E A  S +  ++ + E+E+   ID+ K+ GD+IGG++E IV+GVP GLGS+V +D+K
Sbjct: 181 NLKELAEASPVRCLDKEAEQEMIAAIDQAKENGDSIGGVVEVIVEGVPIGLGSHVHYDRK 240

Query: 239 LDAKLAQAVLSINAFKGVEFGAGFDMGFQKGSQVMDEITWTPTQGYGRQTNHLGGFEGGM 298
           LDAK+A AV+SINAFKGVEFG GF+   + GS+V DEI W   +GY R++N+LGGFEGGM
Sbjct: 241 LDAKIAAAVMSINAFKGVEFGIGFEAASKPGSEVHDEIAWDEERGYYRKSNNLGGFEGGM 300

Query: 299 TTGQPLVVKGVMKPIPTLYKPLMSVDIDSHEPYKATVERSDPTALPAAGVIMENVVATVL 358
           T G P+VV+GVMKPIPTLYKPL SVDI + EP+ A++ERSD  A+PAA V+ E VVA  +
Sbjct: 301 TNGMPIVVRGVMKPIPTLYKPLQSVDIATKEPFAASIERSDSCAVPAAAVVAEAVVAWEV 360

Query: 359 AKEILETFSSTTMSELQKAFSDYRAYVKQF 388
           A  +LE F +  + E++K   ++    + F
Sbjct: 361 ANALLERFGADQVEEIEKNIREFNEKARLF 390
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 284/388 (73%), Positives = 333/388 (85%)

Query: 1   MRYLTAGESHGPSLTAIIEGIPAGLKLSAKDINEDLKRRQGGYGRGNRMKIETDQVIISS 60
           +RYLTAGESHGPSLTAIIEGIPAGL L   DI+ +L+RRQGGYGRG RM IETD+V ISS
Sbjct: 1   LRYLTAGESHGPSLTAIIEGIPAGLTLHPADIDHELQRRQGGYGRGARMSIETDRVQISS 60
```

```
Query: 61  GVRHGKTLGSPITLTVTNKDHSKWLDIMSVEDIEERLKQKRRIKHPRPGHADLVGGIKYR 120
           GVRHGKT G+PITLTV NKDH KWLD+M+V DIEE LK KRR+KHPRPGHADLVGGIKY
Sbjct: 61  GVRHGKTTGAPITLTVINKDHQKWLDVMAVGDIEETLKLKRRVKHPRPGHADLVGGIKYH 120

Query: 121 FDDLRNALERSSARETTMRVAIGAIAKRILKEIGIEIANHIVVFGGKEITVPDKLTVQQI 180
           F+DLR+ALERSSARETTMRVA+GA+AKRIL E+GI++ +HI++FGG  IT+P KL+ +++
Sbjct: 121 FNDLRDALERSSARETTMRVAVGAVAKRILAELGIDMLHHILIFGGITITIPSKLSFREL 180

Query: 181 KVLSSQSQVAIVNPSFEQEIKDYIDSVKKAGDTIGGVVETIVGGVPVGLGSYVHWDRKLD 240
           +  +  S+++IVNP  E+EIK YID +KK GDTIGG++ETIV GVP GLGSYV WD+KLD
Sbjct: 181 QERALHSELSIVNPKQEEEIKTYIDKIKKEGDTIGGIIETIVQGVPAGLGSYVQWDKKLD 240

Query: 241 AKIAQAVVSINAFKGVEFGLGFKSGFLKGSQVMDSISWTKDQGYIRQSNNLGGFEGGMTN 300
           AK+AQAV+SINAFKGVEFG GF  GF KGSQVMD I+WT  QGY RQ+N+LGGFEGGMT
Sbjct: 241 AKLAQAVLSINAFKGVEFGAGFDMGFQKGSQVMDEITWTPTQGYGRQTNHLGGFEGGMTT 300

Query: 301 GEPIIVRGVMKPIPTLYKPLMSVDIDTHEPYRATVERSDPTALPAAGVVMEAVVATVLVT 360
           G+P++V+GVMKPIPTLYKPLMSVDID+HEPY+ATVERSDPTALPAAGV+ME VVATVL
Sbjct: 301 GQPLVVKGVMKPIPTLYKPLMSVDIDSHEPYKATVERSDPTALPAAGVIMENVVATVLAK 360

Query: 361 EVLEKFSSDNMYELKEAVKLYRNYVDHF 388
           E+LE FSS  M EL++A   YR YV  F
Sbjct: 361 EILETFSSTTMSELQKAFSDYRAYVKQF 388
```

A related GBS gene <SEQ ID 8617> and protein <SEQ ID 8618> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 9
McG: Discrim Score:      -2.42
GvH: Signal Score (-7.5): -3.23
     Possible site: 15
>>> Seems to have no N-terminal signal sequence
ALOM program   count: 1 value:  -4.67 threshold:  0.0
   INTEGRAL    Likelihood = -4.67   Transmembrane  343 - 359 ( 341 - 364)
   PERIPHERAL  Likelihood =  0.69     214
 modified ALOM score:   1.43

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.2869(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
57.7/73.8% over 354aa
                                                             Bacillus subtilis
  EGAD|20299| chorismate synthase Insert characterized
  SP|P31104|AROC_BACSU CHORISMATE SYNTHASE (EC 4.6.1.4) (5-ENOLPYRUVYLSHIKIMATE-3-PHOSPHATE
PHOSPHOLYASE)
 (VEGETATIVE PROTEIN 216) (VEG216). Edit characterized
  GP|143806|gb|AAA20859.1||M80245 AroF Insert characterized
  GP|2634689|emb|CAB14187.1||Z99115 chorismate synthase Insert characterized
  PIR|C69590|C69590 chorismate synthase aroF - Insert characterized

ORF00121(301 - 1359 of 1719)
EGAD|20299|BS2267(1     -     355     of     368)     chorismate     synthase     {Bacillus
subtilis}SP|P31104|AROC_BACSU CHORISMATE SYNTHASE (EC 4.6.1.4) (5-ENOLPYRUVYLSHIKIMATE-3-
PHOSPHATE PHOSPHOLYASE) (VEGETATIVE PROTEIN 216) (VEG216).GP|143806|gb|AAA20859.1||M80245
AroF  {Bacillus  subtilis}GP|2634689|emb|CAB14187.1||Z99115  chorismate  synthase  {Bacillus
subtilis}PIR|C69590|C69590 chorismate synthase aroF - Bacillus subtilis
%Match = 35.0
%Identity = 57.6  %Similarity = 73.7
Matches = 204  Mismatches = 92  Conservative Sub.s = 57

75        105        135        165        195        225        255        285
```

```
IQLSRVAERKNLMPRGISQDIYNMCLKFGLPVHYAEWDKDVLFDILSHDKKASGQFIKIVILPQLGSATVHQIPLEEMRD

315       345       375       405       435       465       495       525
YLEK*MRYLTAGESHGPSLTAIIEGIPAGLKLSAKDINEDLKRRQGGYGRGNRMKIETDQVIISSGVRHGKTLGSPITLT
     |||||||||||| || ||||:|||| :: :||| :| ||| |:||| ||:|| ||  | ||||| :|||||| |
     MRYLTAGESHGPQLTTIIEGVPAGLYITEEDINFELARRQKGHGRGRRMQIEKDQAKIMSGVRHARTLGSPIALV
             10        20        30        40        50        60        70

555                 609       639       669       699       729       759
VTNKDHSKWLDIMSVEDI--EERLKQKRRIKHPRPGHADLVGGIKYRFDDLRNALERSSARETTMRVAIGAIAKRILKEI
| | |   |  ||    |  :|  : ||:|  |||||||| | |||   |:|| ||||||||||:||| ||:||:|| |:
VENNDWKHWTKIMGAAPITEDEEKEMKRQISRPRPGHADLNGAIKYNHRDMRNVLERSSARETTVRVAAGAVAKKILSEL
             90       100       110       120       130       140       150

789       819       849       879       909       939       969       999
GIEIANHIVVFGGKEITVPDKLTVQQIKVLSSQSQVAIVNPSFEQEIKDYIDSVKKAGDTIGGVVETIVGGVPVGLGSYV
||::| |::  |  :         ::: :: :: :| |   :    :::   ||  |  ||:|||:|| || |:|||:||||
GIKVAGHVLQIGAVKAEKTGYTSIEDLQRVTEESPVRCYDEEAGKKMMAAIDEAKANGDSIGGIVEVIVEGMPVGVGSYV
            170       180       190       200       210       220       230

1029      1059      1089      1119      1149      1179      1209      1239
HWDRKLDAKIAQAVVSINAFKGVEFGLGFKSGFLKGSQVMDSISWTKDQGYIRQSNNLGGFEGGMTNGEPIIVRGVMKPI
|:|||||:|:| ||:|||||||||||:||::    ||:| | | | :::|| | :| |||:||||| | ||:||||||||
HYDRKLDSKLAAAVLSINAFKGVEFGIGFEAAGRNGSEVHDEIIWDEEKGYTRATNRLGGLEGGMTTGMPIVVRGVMKPI
            250       260       270       280       290       300       310

1269      1299      1329      1359      1389      1419      1449      1479
PTLYKPLMSVDIDTHEPYRATVERSDPTALPAAGVVMEAVVATVLVTEVLEKFSSDNMYN*KKL*NYIAIMLIIFK*KLV
||||||| ||||:| ||: |::||||  |:||| || ||:    |   :
PTLYKPLKSVDIETKEPFSASIERSDSCAVPAASVVAEALSLGKLQPSLNNSD
            330       340       350       360
```

SEQ ID 8618 (GBS192) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 18 (lane 4; MW 44kDa).

GBS192-His was purified as shown in Figure 196, lane 4.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 608

A DNA sequence (GBSx0648) was identified in *S.agalactiae* <SEQ ID 1885> which encodes the amino acid sequence <SEQ ID 1886>. This protein is predicted to be 3-dehydroquinate synthase (aroB). Analysis of this protein sequence reveals the following:

```
Possible site: 24
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -3.82   Transmembrane   99 - 115 (  98 - 116)

----- Final Results -----
               bacterial membrane --- Certainty=0.2529(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAA18068 GB:D90911 3-dehydroquinate synthase [Synechocystis sp.]
 Identities = 138/351 (39%), Positives = 200/351 (56%), Gaps = 4/351 (1%)

Query: 3    VEVDLPNHPYHIKIEEGCFSEAGDWVSHLWQKQMITIITDSNVEILYGESLVNQLKKQGF 62
            + V LP  PY ++I  G  +   D ++ L   + I ++++  +   YGE ++  L++ G+
Sbjct: 5    IPVPLPQSPYQVQIVPGGLAAIADHLAPLGLGKKIMVVSNPEIYDYYGEVVIQALQRAGY 64
```

```
Query: 63  TVHVFSFAAGEASKTLEVANRIYAFLAKHHMTRSDGIIALGGGVVGDLAAFVASTYMRGI 122
            V      AGE  KTL   N +Y    + ++ R+  +++LGGGV+GD+  F A+T++RGI
Sbjct: 65  EVFQHLIPAGETHKTLASINELYDVAFQANLERNSTLLSLGGGVIGDMTGFGAATWLRGI 124

Query: 123 HFLQIPTSLTAQVDSSIGGKTGVNTSFAKNMVGTFAQPDGVLIDPVTLKTLGNRELVEGM 182
           +F+Q+PTSL A VD+SIGGKTGVN    KN++G F QP  V IDPV LKTL  RE   GM
Sbjct: 125 NFVQVPTSLLAMVDASIGGKTGVNHPQGKNLIGAFYQPRLVYIDPVVLKTLPEREFRAGM 184

Query: 183 GEVIKYGLIDDIKLWHILEEMD--GTIDSILDNALA-IIYHSCQVKRKHVLADQYDKGLR 239
            EVIKYG+I D +L+  LEE +   +ID + D  L  II  SCQ K   V  D+ + GLR
Sbjct: 185 AEVIKYGVIWDSELFTALEEAEDLSSIDRLPDELLTKIIQRSCQAKVDVVSQDEKEAGLR 244

Query: 240 MHLNFGHTIGHAIEVHAGYGEIMHGEAVAIGMIQLSRVAERKNLMPRGISQDIYNMCLKF 299
             LN+GHT+GH +E   GYG I HGEAVAIGM   +++A    L  + +      + LK
Sbjct: 245 AILNYGHTVGHGVESLTGYGVINHGEAVAIGMEAAAKIAHYLGLCDQSLGDRQRQLLLKT 304

Query: 300 GLPVHY-AEWDKDVLFDILSHDKKASGQFIKIVILPQLGSATVHQIPLEEM 349
            LP         + L   L HDKK     ++ ++   +G  T+     +E+
Sbjct: 305 KLPTEMPPTLAVENLLASLLHDKKVKAGKVRFILPTAIGQVTISDAVTDEV 355
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1887> which encodes the amino acid sequence <SEQ ID 1888>. Analysis of this protein sequence reveals the following:

```
        Possible site: 60
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.43    Transmembrane   97 - 113 (  97 - 114)

----- Final Results -----
               bacterial membrane --- Certainty=0.1171(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:BAA18068 GB:D90911 3-dehydroquinate synthase [Synechocystis sp.]
 Identities = 123/349 (35%), Positives = 190/349 (54%), Gaps = 9/349 (2%)

Query: 1   MPQTLHVHSRVKDYDILFTDHVLKTLADCLGERKQ-RKLLFITDQTVYHLYQTLFEEFAQ 59
           M  T+ V      Y +      L  +AD L      +K++ +++  +Y  Y  +  +  Q
Sbjct: 1   MATTIPVPLPQSPYQVQIVPGGLAAIADHLAPLGLGKKIMVVSNPEIYDYYGEVVIQALQ 60

Query: 60  Q--YNAFVHVCPPGGQSKSLERVSAIYDQLIAENFSKKDMIVTIGGGVVGDLGGFVAATY 117
           +  Y  F H+ P G   K+L  ++ +YD     N  +   ++++GGGV+GD+ GF AAT+
Sbjct: 61  RAGYEVFQHLIPAGETHKTLASINELYDVAFQANLERNSTLLSLGGGVIGDMTGFGAATW 120

Query: 118 YRGIPYIQIPTTLLSQVDSSIGGKVGVHFKGLTNMIGSIYPPEAIIISTTFLETLPQREF 177
            RGI ++Q+PT+LL+ VD+SIGGK GV+     N+IG+ Y P  + I    L+TLP+REF
Sbjct: 121 LRGINFVQVPTSLLAMVDASIGGKTGVNHPQGKNLIGAFYQPRLVYIDPVVLKTLPEREF 180

Query: 178 SCGISEMLKIGFIHDRPLFQQLRDFQ-----KETDKQGLERLIYQSISNKKRIVEQDEFE 232
             G++E++K G I D  LF  L + +          + L ++I +S   K  +V QDE E
Sbjct: 181 RAGMAEVIKYGVIWDSELFTALEEAEDLSSIDRLPDELLTKIIQRSCQAKVDVVSQDEKE 240

Query: 233 NGLRMSLNFGHTLGHAIESLCHHDFYHHGEAIAIGMVVDAKLAVSKGLLPKEDLDSLLQV 292
            GLR  LN+GHT+GH +ESL  +   +HGEA+AIGM   AK+A   GL  +   D   Q+
Sbjct: 241 AGLRAILNYGHTVGHGVESLTGYGVINHGEAVAIGMEAAAKIAHYLGLCDQSLGDRQRQL 300

Query: 293 FERYQLPTTLERADVSATSLFDVFKTDKKNSEQHIIFILPTETGFTTLA 341
             + +LPT +    ++  +L      DKK     + FILPT  G  T++
Sbjct: 301 LLKTKLPTEMP-PTLAVENLLASLLHDKKVKAGKVRFILPTAIGQVTIS 348
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 121/332 (36%), Positives = 182/332 (54%), Gaps = 7/332 (2%)

Query: 12  YHIKIEEGCFSEAGDWVSHLWQKQMITIITDSNVEILYGESLVNQLKKQGFTVHVFSFAA 71
           Y I   +       D +    Q++++  ITD  V  LY ++L  +  +Q +   V
```

```
Sbjct: 14  YDILFTDHVLKTLADCLGERKQRKLL-FITDQTVYHLY-QTLFEEFAQQ-YNAFVHVCPP 70

Query: 72  GEASKTLEVANRIYAFLAKHHMTRSDGIIALGGGVVGDLAAFVASTYMRGIHFLQIPTSL 131
            G  SK+LE  + IY  L   + ++ D I+ +GGGVVGDL  FVA+TY RGI ++QIPT+L
Sbjct: 71  GGQSKSLERVSAIYDQLIAENFSKKDMIVTIGGGVVGDLGGFVAATYYRGIPYIQIPTTL 130

Query: 132 TAQVDSSIGGKTGVNTSFAKNMVGTFAQPDGVLIDPVTLKTLGNRELVEGMGEVIKYGLI 191
             +QVDSSIGGK GV+     NM+G+   P+ ++I    L+TL  RE   G+ E++K G I
Sbjct: 131 LSQVDSSIGGKVGVHFKGLTNMIGSIYPPEAIIISTTFLETLPQREFSCGISEMLKIGFI 190

Query: 192 DDIKLWHILEEMDGTIDSILDNALAIIYHSCQVKRKHVLADQYDKGLRMHLNFGHTIGHA 251
             D  L+  L +     D         +IY S   K++ V  D+++ GLRM LNFGHT+GHA
Sbjct: 191 HDRPLFQQLRDFQKETDK--QGLERLIYQSISNKKRIVEQDEFENGLRMSLNFGHTLGHA 248

Query: 252 IEVHAGYGEIMHGEAVAIGMIQLSRVAERKNLMPRGISQDIYNMCLKFGLP--VHYAEWD 309
            IE    +    HGEA+AIGM+  +++A  K L+P+     +  +  ++ LP  +  A+
Sbjct: 249 IESLCHHDFYHHGEAIAIGMVVDAKLAVSKGLLPKEDLDSLLQVFERYQLPTTLERADVS 308

Query: 310 KDVLFDILSHDKKASGQFIKIVILPQLGSATV 341
               LFD+   DKK S Q I  ++  + G  T+
Sbjct: 309 ATSLFDVFKTDKKNSEQHIIFILPTETGFTTL 340
```

SEQ ID 1886 (GBS336) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 62 (lane 2; MW 42.7kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 67 (lane 5; MW 68kDa).

The GBS336-GST fusion product was purified (Figure 209, lane 4) and used to immunise mice. The resulting antiserum was used for FACS (Figure 310), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 609

A DNA sequence (GBSx0649) was identified in *S.agalactiae* <SEQ ID 1889> which encodes the amino acid sequence <SEQ ID 1890>. Analysis of this protein sequence reveals the following:

```
Possible site: 47
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3884(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9973> which encodes amino acid sequence <SEQ ID 9974> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB14240 GB:Z99116 3-dehydroquinate dehydratase [Bacillus subtilis]
 Identities = 70/233 (30%), Positives = 127/233 (54%), Gaps = 12/233 (5%)

Query: 2   KIVVPVMPRSLEEA-QEIDLSKFDSVDIIEWRADALPK----DDIINVAPAIFEKFAGHE 56
            KI++P+M ++ ++   E +  K  + DI+EWR D   K    + +  +   + +
Sbjct: 17  KIIIPLMGKTEKQILNEAEAVKLLNPDIVEWRVDVFEKANDREAVTKLISKLRKSLEDKL 76

Query: 57  IIFTLRTTREGGNIVLSDAEYVELIQKINSIYNPDYIDFEYFSHKEVFQEMLEFPN---- 112
             +FT RT +EGG++ + ++ Y+ L++      + D ID E FS     + ++
Sbjct: 77  FLFTFRTHKEGGSMEMDESSYLALLESAIQTKDIDLIDIELFSGDANVKALVSLAEENNV 136
```

```
Query: 113 -LVLSYHNFQETP--ENIMEIFSELTALAPRVVKIAVMPKNEQDVLDVMNYTRGFKTINP 169
            +V+S H+F++TP  + I+    ++  L   + K+AVMP +  D+L +++ T   KTI
Sbjct: 137 YVVMSNHDFEKTPVKDEIISRLRKMQDLGAHIPKMAVMPNDTGDLLTLLDATYTMKTIYA 196

Query: 170 DQVYATVSMSKIGRISRFAGDVTGSSWTFAYLDSSIAPGQITISEMKRVKALL 222
           D+   T+SM+  G ISR +G+V GS+ TF   + + APGQI +SE++ V  +L
Sbjct: 197 DRPIITMSMAATGLISRLSGEVFGSACTFGAGEEASAPGQIPVSELRSVLDIL 249
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1891> which encodes the amino acid sequence <SEQ ID 1892>. Analysis of this protein sequence reveals the following:

```
Possible site: 17
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3248(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 160/225 (71%), Positives = 198/225 (87%)

Query: 1   MKIVVPVMPRSLEEAQEIDLSKFDSVDIIEWRADALPKDDIINVAPAIFEKFAGHEIIFT 60
           M+IV PVMPR  +EAQ ID+SK++ V++IEWRAD LPKD+I+ VAPAIFEKFAG EIIFT
Sbjct: 1   MRIVAPVMPRHFDEAQAIDISKYEDVNLIEWRADFLPKDEIVAVAPAIFEKFAGKEIIFT 60

Query: 61  LRTTREGGNIVLSDAEYVELIQKINSIYNPDYIDFEYFSHKEVFQEMLEFPNLVLSYHNF 120
           LRT +EGGNI LS  EYV++I++IN+IYNPDYIDFEYF+HK VFQEML+FPNL+LSYHNF
Sbjct: 61  LRTVQEGGNITLSSQEYVDIIKEINAIYNPDYIDFEYFTHKSVFQEMLDFPNLILSYHNF 120

Query: 121 QETPENIMEIFSELTALAPRVVKIAVMPKNEQDVLDVMNYTRGFKTINPDQVYATVSMSK 180
           +ETPEN+ME FSE+T LAPRVVKIAVMP++EQDVLD+MNYTRGFKT+NP+Q +AT+SM K
Sbjct: 121 EETPENLMEAFSEMTKLAPRVVKIAVMPQSEQDVLDLMNYTRGFKTLNPEQEFATISMGK 180

Query: 181 IGRISRFAGDVTGSSWTFAYLDSSIAPGQITISEMKRVKALLDAD 225
           +GR+SRFAGDV GSSWT+  LD    PGQ+T+++MKR+  +L+ D
Sbjct: 181 LGRLSRFAGDVIGSSWTYVSLDHVSGPGQVTLNDMKRIIEVLEMD 225
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 610

A DNA sequence (GBSx0650) was identified in *S.agalactiae* <SEQ ID 1893> which encodes the amino acid sequence <SEQ ID 1894>. Analysis of this protein sequence reveals the following:

```
Possible site: 17
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1195(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 611

A DNA sequence (GBSx0651) was identified in *S.agalactiae* <SEQ ID 1895> which encodes the amino acid sequence <SEQ ID 1896>. Analysis of this protein sequence reveals the following:

```
Possible site: 41
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3431(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB15862 GB:Z99123 alternate gene name: ipa-19d~similar to
           hypothetical proteins [Bacillus subtilis]
 Identities = 161/396 (40%), Positives = 235/396 (58%), Gaps = 11/396 (2%)

Query: 1   MNKLKVNSVVERKIKSGAQLLEKKDFDTSLVNQ----LVQLFSQSN-QFLGMAYLSPQNK 55
           M  L +      KIK G  L+EK+    S  +     LV + S+S  +FL   Y   QNK
Sbjct: 1   MKLLTLKKAHAAKIKKGYPLIEKEALAGSAGHMKEGDLVDIVSESGGEFLARGYYGLQNK 60

Query: 56  GIGWLLSRQVFD-FNHDYFVSLFEKSREKRQKFEKSSQTTAYRLFNQDGDNFGGLTIDFY 114
           G+GW L+R   +  +  +F+S   K+ + R K  ++  TTA+RLFN +GD  GG+TID+Y
Sbjct: 61  GVGWTLTRNKHEQIDQAFFLSKLTKAAQARAKLFEAQDTTAFRLFNGEGDGVGGVTIDYY 120

Query: 115 SDYALFSWYNEFVYTNRQMIVAAFKQVYPNIKGAYEKIRFKGLDF---ESAHLYGQEAPE 171
             Y L  WY++ +YT + M+++A  ++  + K  YEK RF        +   + G+
Sbjct: 121 DGYLLIQWYSKGIYTFKDMLISALDEMDLDYKAIYEKKRFDTAGQYVEDDDFVKGRRGEF 180

Query: 172 SFLILENNIKYSVFLNDGLMTGIFLDQHDVRKALATNLSEGKKVLNMFSYTAAFSVAAAV 231
             +I EN I+Y+V LN+G MTGIFLDQ  VRKA+    ++GK VLN FSYT AFSVAAA+
Sbjct: 181 PIIIQENGIQYAVDLNEGAMTGIFLDQRHVRKAIRDRYAKGKTVLNTFSYTGAFSVAAAL 240

Query: 232 GGALETTSVDLAKRSRELSKAHFDANQIVTDNHRFIVMDVFEYYKYAKRKHLSYDVIVID 291
           GGA +TTSVD+A RS   +   F  N++  + H   VMDVF Y+ YA +K L +D+I++D
Sbjct: 241 GGAEKTTSVDVANRSLAKTIEQFSVNKLDYEAHDIKVMDVFNYFSYAAKKDLRFDLIILD 300

Query: 292 PPSFARNKKQTFSVTKDYYKLIEQALDILTPGGTIIASTNAANLTVSQFKKQLEKGFGKA 351
           PPSFAR KK+TFS  KDY  L+++ + I    G I+ASTN++   + +FK  ++  F +
Sbjct: 301 PPSFARTKKRTFSAAKDYKNLLKETIAITADKGVIVASTNSSAFGMKKFKGFIDAAFKET 360

Query: 352 SHNYISLQQ--LPEDFTINDKDQQSNYLKVFTIKVK 385
           +  Y  +++  LPEDF       + NYLKV  ++ K
Sbjct: 361 NERYTIIEEFTLPEDFKTISAFPEGNYLKVVLLQKK 396
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1897> which encodes the amino acid sequence <SEQ ID 1898>. Analysis of this protein sequence reveals the following:

```
Possible site: 29
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2699(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 259/386 (67%), Positives = 315/386 (81%), Gaps = 1/386 (0%)

Query: 1   MNKLKVNSVVERKIKSGAQLLEKKDFDT-SLVNQLVQLFSQSNQFLGMAYLSPQNKGIGW 59
           MNKL ++S VE+K+ +G QLL++KDF      NQLVQL ++SN+ +G AY+S QNKGIGW
Sbjct: 1   MNKLYIDSFVEKKLTAGVQLLDEKDFSNIKEKNQLVQLVTKSNRPIGTAYISKQNKGIGW 60
```

```
Query: 60  LLSRQVFDFNHDYFVSLFEKSREKRQKFEKSSQTTAYRLFNQDGDNFGGLTIDFYSDYAL 119
            L  +  D +  YFVSLF  ++ KRQ F +S +T AYRLFNQ+GD FGG+TID Y D+A+
Sbjct: 61  YLGPEKIDLSISYFVSLFSVAKAKRQDFAQSDETNAYRLFNQEGDGFGGVTIDLYKDFAV 120

Query: 120 FSWYNEFVYTNRQMIVAAFKQVYPNIKGAYEKIRFKGLDFESAHLYGQEAPESFLILENN 179
           FSWYN FVY  ++MI+ AF+QV+P +KGAYEK RFKG D E+AHLYG+ A E+F ILEN
Sbjct: 121 FSWYNAFVYDKKEMIMEAFQQVFPEVKGAYEKCRFKGPDTETAHLYGELAQETFSILENG 180

Query: 180 IKYSVFLNDGLMTGIFLDQHDVRKALATNLSEGKKVLNMFSYTAAFSVAAAVGGALETTS 239
           I Y VFLN+GLMTGIFLDQHDVR+AL   L+ GK +LN+FSYTAAFSVAAA+GGA+ETTS
Sbjct: 181 IAYQVFLNEGLMTGIFLDQHDVRRALVDGLAMGKSLLNLFSYTAAFSVAAAMGGAIETTS 240

Query: 240 VDLAKRSRELSKAHFDANQIVTDNHRFIVMDVFEYYKYAKRKHLSYDVIVIDPPSFARNK 299
           VDLAKRSRELS AHF+ NQ+   +H F+VMDVFEY+KYAKRK L +DVIVIDPPSFARNK
Sbjct: 241 VDLAKRSRELSLAHFEHNQLNLASHHFVVMDVFEYFKYAKRKKLIFDVIVIDPPSFARNK 300

Query: 300 KQTFSVTKDYYKLIEQALDILTPGGTIIASTNAANLTVSQFKKQLEKGFGKASHNYISLQ 359
           KQTFSV++DY+KLI +ALDIL+P GTIIASTNAAN+TVSQFKKQ+ KGFG      ++LQ
Sbjct: 301 KQTFSVSRDYHKLITEALDILSPKGTIIASTNAANMTVSQFKKQIIKGFGSRRPESMTLQ 360

Query: 360 QLPEDFTINDKDQQSNYLKVFTIKVK 385
           QLP DFTIN  D++SNYLKVFTIKV+
Sbjct: 361 QLPSDFTINKADERSNYLKVFTIKVR 386
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 612

A DNA sequence (GBSx0652) was identified in *S.agalactiae* <SEQ ID 1899> which encodes the amino acid sequence <SEQ ID 1900>. This protein is predicted to be minimal change nephritis transmembrane glycoprotein. Analysis of this protein sequence reveals the following:

```
Possible site: 30
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -6.85   Transmembrane  129 - 145 ( 126 - 152)
   INTEGRAL    Likelihood = -4.88   Transmembrane   48 -  64 (  46 -  69)
   INTEGRAL    Likelihood = -4.83   Transmembrane   75 -  91 (  74 -  97)
   INTEGRAL    Likelihood = -4.62   Transmembrane   16 -  32 (  15 -  34)
   INTEGRAL    Likelihood = -2.28   Transmembrane  163 - 179 ( 163 - 182)

----- Final Results -----
               bacterial membrane --- Certainty=0.3739(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB12545 GB:Z99107 alternate gene name: yetP~similar to
           hypothetical proteins [Bacillus subtilis]
 Identities = 299/676 (44%), Positives = 415/676 (61%), Gaps = 33/676 (4%)

Query: 2   KKIKDFASRAINTRLGFILLLVVIYWLKTIWAYHTDFNLGLENSYQLFLTIINPIPLGLL 61
           KK++    +  + +L F +L V+++W KT  +Y T+FNLG++ + Q  L I NP    +
Sbjct: 9   KKVEVAMKKLFSYKLSFFVLAVILFWAKTYLSYKTEFNLGVKGTTQEILLIFNPFSSAVF 68

Query: 62  IIGLALYVKRTKAFYITAFITYAIVNILLIANAIYYREFSDFITVSAVLASSKTSAGLGD 121
            +GLAL  K  K+  I   I + ++  +L AN ++YR F DF+T   +  S      +GD
Sbjct: 69  FLGLALLAKGRKSAIIMLIIDF-LMTFVLYANILFYRFFDDFLTFPNIKQSGNVG-NMGD 126

Query: 122 SALNLLRIWDLVYVFDFIILIFLFATKKIHLDDRPFNKRASFSITALSGL-LFSINLFLA 180
              +++   D+ Y  D IILI +   +   L +    KR + S+  LSG+ LF INL  A
Sbjct: 127 GIFSIMAGHDIFYFLDIIILIAVLIWRP-ELKEYKMKKRFA-SLVILSGIALFFINLHYA 184

Query: 181 EIDRPELLSRGFSNTYIVKALGLPSFSIYSGNQTYQAQKERNGATAQELATAKKYVAEHY 240
```

```
           E DRP+LL+R F   YIVK LGL +++IY G QT Q + +R  A++ +L + + Y   HY
Sbjct: 185 EKDRPQLLTRTFDRNYIVKYLGLYNYTIYDGVQTAQTETQRAYASSDDLTSVENYTTSHY 244

Query: 241 AKPNPEYYGIGKGRNVIMIHLESFQQFLIDYKLNIDGKEHVVTPFINSLYHSKETVS-FS 299
           AKPN EY+G  KG+N+I IHLESFQ FLIDYKLN  G+E  VTPF+N L H  E V+ F
Sbjct: 245 AKPNAEYFGSAKGKNIIKIHLESFQSFLIDYKLN--GEE--VTPFLNKLAHGGEDVTYFD 300

Query: 300 NFFHQVKAGKTSDAETLMENSLFGLSSGSFMVNYGGENTQFAAPHILAQNGGYSSAVFHG 359
           NFFHQ   GKTSDAE  M+NS+FGL  GS  V   GENT  + P IL Q  GY+SAV HG
Sbjct: 301 NFFHQTGQGKTSDAELTMDNSIFGLPEGSAFVT-KGENTYQSLPAILDQKEGYTSAVLHG 359

Query: 360 NVGTFWNRNNAYKQWGYDYFFDSSYFSKQTKDNSFQYGLNDKYMFADSIKYLEHMQQPFY 419
           +  +FWNR+  YK  GYD FFD+S +   + +N    GL DK  F +SI  LE ++QPFY
Sbjct: 360 DYKSFWNRDQIYKHIGYDKFFDASTYD-MSDENVINMGLKDKPFFTESIPKLESLKQPFY 418

Query: 420 TKFITVSNHYPYTSLKGESDEEGFPLAKTNDETINGYFATANYLDTALKSFFEYLKAAGV 479
              IT++NHYP+     +  +      A T D T++ YF TA YLD AL+ FF+ LK AG+
Sbjct: 419 AHLITLTNHYPFNL---DEKDASLKKATTGDNTVDSYFQTARYLDEALEQFFKELKEAGL 475

Query: 480 YDNSIIVMYGDHYGISNTRNPSLAELLGKDPETWSEYDNAMLQRVPYMIHIPGYSKGFIS 539
           YDNS+I++YGDH GIS   N ++ E+LGK+    ++Y NA  QRVP MI +PG  KG ++
Sbjct: 476 YDNSVIMIYGDHNGISENHNRAMKEILGKE---ITDYQNAQNQRVPLMIRVPG-KKGGVN 531

Query: 540 NTYGGEVDNLPTLLHILGIDTSKYTQLGQDLLSKDNKQMVAMRTTGQYITPKYTNYSGHL 599
           +TYGGE+D +PTLLH+ GID+ KY   G DL SKD+   VA R  G ++TPKYT+    +
Sbjct: 532 HTYGGEIDVMPTLLHLEGIDSQKYINFGTDLFSKDHDDTVAFR-NGDFVTPKYTSVDNII 590

Query: 600 YYTDSGQEITNPDETTKAEIKAIRDATNKQLSTSDSIQTGDLLRFDENNGLKTVEVEKFN 659
           Y T +G+++   +ET     K ++   N+QLS SDS+   DLLRF + N  K V+   ++
Sbjct: 591 YDTKTGEKLKANEET-----KNLKTRVNQQLSLSDSVLYKDLLRFHKLNDFKAVDPSDYH 645

Query: 660 YTHSLKALKAKERKLK 675
           Y         KE+++K
Sbjct: 646 Y--------GKEKEIK 653
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1901> which encodes the amino acid sequence <SEQ ID 1902>. Analysis of this protein sequence reveals the following:

```
Possible site: 48
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -6.85   Transmembrane   90 - 106 (  88 - 112)
   INTEGRAL    Likelihood = -5.68   Transmembrane  146 - 162 ( 139 - 165)
   INTEGRAL    Likelihood = -4.99   Transmembrane   63 -  79 (  60 -  84)
   INTEGRAL    Likelihood = -3.98   Transmembrane  178 - 194 ( 176 - 197)
   INTEGRAL    Likelihood = -0.59   Transmembrane   31 -  47 (  31 -  47)

----- Final Results -----
               bacterial membrane --- Certainty=0.3739(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 533/713 (74%), Positives = 603/713 (83%)

Query: 1   MKKIKDFASRAINTRLGFILLLVVIYWLKTIWAYHTDFNLGLENSYQLFLTIINPIPLGL 60
           +KK K   +  INTRLGFI+ L+  YW+KT+WAYHTDF+L L N YQ+FLTIINPIPL
Sbjct: 16  VKKFKTLITGFINTRLGFIITLLFCYWIKTLWAYHTDFSLDLGNIYQVFLTIINPIPLAF 75

Query: 61  LIIGLALYVKRTKAFYITAFITYAIVNILLIANAIYYREFSDFITVSAVLASSKTSAGLG 120
           L++G+ALYVK T+AFYI +++ Y I+NILLI+N+IYYREFSDFITVSA+LASSK SAGLG
Sbjct: 76  LLLGVALYVKNTRAFYICSWVVYIILNILLISNSIYYREFSDFITVSAMLASSKVSAGLG 135

Query: 121 DSALNLLRIWDLVYVFDFIILIFLFATKKIHLDDRPFNKRASFSITALSGLLFSINLFLA 180
           DSALNLLRIWD++Y+ DFIILI L   KKI  D RPFNKRA+F+ITALS LL SINLFLA
Sbjct: 136 DSALNLLRIWDIIYILDFIILISLSIAKKIKNDQRPFNKRAAFAITALSSLLLSINLFLA 195
```

```
Query: 181 EIDRPELLSRGFSNTYIVKALGLPSFSIYSGNQTYQAQKERNGATAQELATAKKYVAEHY 240
           EIDRPELL+RGFSNTYIV+ALGLP+F++YSGNQTYQAQKERNGATA+EL   K YV  HY
Sbjct: 196 EIDRPELLTRGFSNTYIVRALGLPAFTLYSGNQTYQAQKERNGATAEELIDVKTYVKGHY 255

Query: 241 AKPNPEYYGIGKGRNVIMIHLESFQQFLIDYKLNIDGKEHVVTPFINSLYHSKETVSFSN 300
           A P+P+Y+GIGKG+N+I++HLESFQQFLIDYKL    KE+ VTPFINSLYHS  T++F N
Sbjct: 256 AAPDPQYFGIGKGKNIIVLHLESFQQFLIDYKLKEGDKEYEVTPFINSLYHSNATLAFPN 315

Query: 301 FFHQVKAGKTSDAETLMENSLFGLSSGSFMVNYGGENTQFAAPHILAQNGGYSSAVFHGN 360
           FFHQVKAGKTSDAET+MENSLFGL+SGSFMVNYGGENTQFA P ILAQ GGY+SAVFHGN
Sbjct: 316 FFHQVKAGKTSDAETMMENSLFGLNSGSFMVNYGGENTQFATPSILAQKGGYTSAVFHGN 375

Query: 361 VGTFWNRNNAYKQWGYDYFFDSSYFSKQTKDNSFQYGLNDKYMFADSIKYLEHMQQPFYT 420
           VGTFWNRNNAYKQWGY+YFFDSSYFSKQ   NSFQYGLNDKYMF DSIKYLE MQQPFYT
Sbjct: 376 VGTFWNRNNAYKQWGYNYFFDSSYFSKQNSKNSFQYGLNDKYMFKDSIKYLEQMQQPFYT 435

Query: 421 KFITVSNHYPYTSLKGESDEEGFPLAKTNDETINGYFATANYLDTALKSFFEYLKAAGVY 480
           KFITVSNHYPYTSLKGES EEGFPLAKT+DETINGYFATANYLD ALKSFF+YLKA G+Y
Sbjct: 436 KFITVSNHYPYTSLKGESSEEGFPLAKTDDETINGYFATANYLDAALKSFFDYLKATGLY 495

Query: 481 DNSIIVMYGDHYGISNTRNPSLAELLGKDPETWSEYDNAMLQRVPYMIHIPGYSKGFISN 540
           DNSI V+YGDHYGISN+RN SLA LLGKD ETWSEYDNAMLQRVPYMIHIPGY+ G I
Sbjct: 496 DNSIFVLYGDHYGISNSRNSSLAPLLGKDSETWSEYDNAMLQRVPYMIHIPGYTNGSIKE 555

Query: 541 TYGGEVDNLPTLLHILGIDTSKYTQLGQDLLSKDNKQMVAMRTTGQYITPKYTNYSGHLY 600
           T+GGE+D LPTLLHILGIDTS++ QLGQDLLS  N Q+VA RT+G Y+TP+YTNYSG LY
Sbjct: 556 TFGGEIDALPTLLHILGIDTSQFVQLGQDLLSPQNSQIVAQRTSGTYMTPEYTNYSGRLY 615

Query: 601 YTDSGQEITNPDETTKAEIKAIRDATNKQLSTSDSIQTGDLLRFDENNGLKTVEVEKFNY 660
            T +G EITNPDE T A+ K IR A  +QL+ SD+IQTGDLLRFD  NGLK ++  +F Y
Sbjct: 616 NTQTGLEITNPDEMTIAKTKEIRSAVAQQLAASDAIQTGDLLRFDTQNGLKAIDPNQFIY 675

Query: 661 THSLKALKAKERKLKDRSTSIYSKHNNKSTVDLFHAPSYLELQDPNKTHKTSK 713
           T  LK LK    KL   STS+YSK+ +KST  LF APSYLEL        TS+
Sbjct: 676 TKQLKQLKDISAKLGSESTSLYSKNGHKSTQKLFKAPSYLELNPVEADAATSE 728
```

A related GBS gene <SEQ ID 8619> and protein <SEQ ID 8620> were also identified. Analysis of this *protein sequence reveals the following:*

```
Lipop Possible site: -1   Crend: 9
McG: Discrim Score:     12.63
GvH: Signal Score (-7.5): -2.99
     Possible site: 30
>>> Seems to have an uncleavable N-term signal seq
ALOM program   count: 5 value:  -6.85 threshold:  0.0
   INTEGRAL    Likelihood = -6.85   Transmembrane  129 - 145 ( 126 - 152)
   INTEGRAL    Likelihood = -4.88   Transmembrane   48 -  64 (  46 -  69)
   INTEGRAL    Likelihood = -4.83   Transmembrane   75 -  91 (  74 -  97)
   INTEGRAL    Likelihood = -4.62   Transmembrane   16 -  32 (  15 -  34)
   INTEGRAL    Likelihood = -2.28   Transmembrane  163 - 179 ( 163 - 182)
   PERIPHERAL  Likelihood =  3.76     103
 modified ALOM score:   1.87

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.3739(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
                    45.2/63.1% over 643aa
                                                                          Bacillus subtilis
  EGAD|107893| hypothetical protein Insert characterized
  GP|2116767|dbj|BAA20118.1||D86418 YfnI Insert characterized
```

GP|2633039|emb|CAB12545.1||Z99107 alternate gene name: yetP~similar to hypothetical proteins Insert characterized

PIR|D69815|D69815 conserved hypothetical protein yfnI - Insert characterized

ORF00125(286 - 2280 of 2742)
EGAD|1078||| S0726(3 - 646 of 653) hypothetical protein { acillus subtilis} GP|2116767|dbj|AA20118.1||D86418 YfnI { acillus subtilis} GP|2633039|emb|CA 12545.1||Z99107 alternate gene name: yetP~similar to hypothetical proteins { acillus subtilis} PIR|D69815|D69815 conserved hypothetical protein yfnI - acillus subtilis
%Match = 28.5
%Identity = 45.1 %Similarity = 63.1
Matches = 297 Mismatches = 227 Conservative Sub.s = 118

```
36        66        96        126       156       186       216       246
FVVKDRPSLRIDLTVKKVEPTG*LNWYQNLFFPVTEHLI*FFFQRQNSL*VYS*TVL*QIFIFFHTEFDLSLPYVTKFYV

276       306       336       366       396       426       456       486
II*SEILSLGKKLKEVPTVKKIKDFASRAINTRLGFILLLVVIYWLKTIWAYHTDFNLGLENSYQLFLTIINPIPLGLLI
             :|:   ||::    :  : :| | :| |:::| ||  :| |:||||:: : |  | | ||    ::
           MNEELKVFKKVEVAMKKLFSYKLSFFVLAVILFWAKTYLSYKTEFNLGVKGTTQEILLIFNPFSSAVFF
                   10        20        30        40        50        60

516       546       576       606       636       666       696       726
IGLALYVKRTKAFYITAFITYAIVNILLIANAIYYREFSDFITVSAVLASSKTSAGLGDSALNLLRIWDLVYVFDFIILI
:||||  |  |:  |   | : ::  :| || ::|| | ||:|   :   |     :||  ::::   |: | :| ||||
LGLALLAKGRKSAIIMLIIDF-LMTFVLYANILFYRFFDDFLTFPNI-KQSGNVGNMGDGIFSIMAGHDIFYFLDIIILI
         80        90        100       110        120       130       140

756       786                 843       873       903       933       963
FLFATKKIHLDDRPFNKRASFSITALSGL-LFSINLFLAEIDRPELLSRGFSNTYIVKALGLPSFSIYSGNQTYQAQKER
      :  | :    ||   |:  |||: || |||  || |||:||:| |   ||||  ||| :::|| | || | : :|
-AVLIWRPELKEYKMKKR-FASLVILSGIALFFINLHYAEKDRPQLLTRTFDRNYIVKYLGLYNYTIYDGVQTAQTETQR
           160        170       180       190       200       210       220

993      1023      1053      1083      1113      1143      1173      1200
NGATAQELATAKKYVAEHYAKPNPEYYGIGKGRNVIMIHLESFQQFLIDYKLNIDGKEHVVTPFINSLYHSKETVS-FSN
  |:: :| : : |   |||||| ||:|  ||:|:| ||||||| ||||||||       ||||:| | |  | |: | |
AYASSDDLTSVENYTTSHYAKPNAEYFGSAKGKNIIKIHLESFQSFLIDYKLN----GEEVTPFLNKLAHGGEDVTYFDN
            240       250       260       270             280       290       300

1230      1260      1290      1320      1350      1380      1410      1440
FFHQVKAGKTSDAETLMENSLFGLSSGSFMVNYGGENTQFAAPHILAQNGGYSSAVFHGNVGTFWNRNNAYKQWGYDYFF
||||   |||||||  |:||:|||  ||  |   ||||  : | || |  ||:|||:||:  :||||:  ||: ||| ||
FFHQTGQGKTSDAELTMDNSIFGLPEGSAFVT-KGENTYQSLPAILDQKEGYTSAVLHGDYKSFWNRDQIYKHIGYDKFF
                320       330        340       350       360       370       380

1470      1500      1530      1560      1590      1620      1650      1680
DSSYFSKQTKDNSFQYGLNDKYMFADSIKYLEHMQQPFYTKFITVSNHYPYTSLKGESDEEGFPLAKTNDETINGYFATA
|:|     : :|    || ||  | :||  || ::||||  :||::||||: :|  | |   :  | | | |:: || ||
DAS-TYDMSDENVINMGLKDKPFFTESIPKLESLKQPFYAHLITLTNHYPF-NLD-EKDAS-LKKATTGDNTVDSYFQTA
        390       400       410       420       430          440       450

1710      1740      1770      1800      1830      1860      1890      1920
NYLDTALKSFFEYLKAAGVYDNSIIVMYGDHYGISNTRNPSLAELLGKDPETWSEYDNAMLQRVPYMIHIPGYSKGFISN
 ||| ||: ||: || ||:||||:|::|||| |||   | :: |:|||     ::| ||  |||| || :||  || :::
RYLDEALEQFFKELKEAGLYDNSVIMIYGDHNGISENHNRAMKEILGK---EITDYQNAQNQRVPLMIRVPG-KKGGVNH
           470       480       490       500          510       520        530

1950      1980      2010      2040      2070      2100      2130      2160
TYGGEVDNLPTLLHILGIDTSKYTQLGQDLLSKDNKQMVAMRTTGQYITPKYTNYSGHLYYTDSGQEITNPDETTKAEIK
|||||:| :|||||: |||: ||  :| ||:|||:   || |  | ::|||||:    :| | :|:::         | |
TYGGEIDVMPTLLHLEGIDSQKYINFGTDLFSKDHDDTVAFR-NGDFVTPKYTSVDNIIYDTKTGEKL-----KANEETK
               550       560       570        580       590            600

2190      2220      2250      2280      2310      2340      2370      2400
AIRDATNKQLSTSDSIQTGDLLRFDENNGLKTVEVEKFNYTHSLKALKAKERKLKDRSTSIYSKHNNKSTVDLFHAPSYL
 ::   |:||| |||:   ||||| : | :| |:   ::|
NLKTRVNQQLSLSDSVLYKDLLRFHKLNDFKAVDPSDYHYGKEKEIK
```

```
              620       630       640       650
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 613

A DNA sequence (GBSx0653) was identified in *S.agalactiae* <SEQ ID 1903> which encodes the amino acid sequence <SEQ ID 1904>. This protein is predicted to be 50S ribosomal protein L20 (rplT). Analysis of this protein sequence reveals the following:

```
Possible site: 37
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3392(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9387> which encodes amino acid sequence <SEQ ID 9388> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB14845 GB:Z99118 ribosomal protein L20 [Bacillus subtilis]
 Identities = 70/89 (78%), Positives = 78/89 (86%)

Query: 1    MFRTAKEQVMNSYYYAYRDRRQKKRDFRKLWITRINAAARMNGLSYSQLMHGLKLAEIEV 60
            +++ A +QVM S  YA+RDRRQKKRDFRKLWITRINAAARMNGLSYS+LMHGLKL+ IEV
Sbjct: 31   LYKVANQQVMKSGNYAFRDRRQKKRDFRKLWITRINAAARMNGLSYSRLMHGLKLSGIEV 90

Query: 61   NRKMLADLAVNDAAAFTALADAAKAKLGK 89
            NRKMLADLAVND  AF  LADAAKA+L K
Sbjct: 91   NRKMLADLAVNDLTAFNQLADAAKAQLNK 119
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1905> which encodes the amino acid sequence <SEQ ID 1906>. Analysis of this protein sequence reveals the following:

```
Possible site: 27
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.06   Transmembrane   94 - 110 (  94 - 110)

----- Final Results -----
               bacterial membrane --- Certainty=0.1022(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 87/89 (97%), Positives = 88/89 (98%)

Query: 1    MFRTAKEQVMNSYYYAYRDRRQKKRDFRKLWITRINAAARMNGLSYSQLMHGLKLAEIEV 60
            +FRTAKEQVMNSYYYAYRDRRQKKRDFRKLWITRINAAARMNGLSYSQLMHGLKLAEIEV
Sbjct: 31   LFRTAKEQVMNSYYYAYRDRRQKKRDFRKLWITRINAAARMNGLSYSQLMHGLKLAEIEV 90

Query: 61   NRKMLADLAVNDAAAFTALADAAKAKLGK 89
            NRKMLADLAV DAAAFTALADAAKAKLGK
Sbjct: 91   NRKMLADLAVADAAAFTALADAAKAKLGK 119
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 614

A DNA sequence (GBSx0654) was identified in *S.agalactiae* <SEQ ID 1907> which encodes the amino acid sequence <SEQ ID 1908>. Analysis of this protein sequence reveals the following:

```
Possible site: 21
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -0.64    Transmembrane   32 -  48 (  32 -  48)
   INTEGRAL    Likelihood = -0.32    Transmembrane    3 -  19 (   3 -  19)

----- Final Results -----
              bacterial membrane --- Certainty=0.1256(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 615

A DNA sequence (GBSx0655) was identified in *S.agalactiae* <SEQ ID 1909> which encodes the amino acid sequence <SEQ ID 1910>. Analysis of this protein sequence reveals the following:

```
Possible site: 33
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood =-12.63    Transmembrane  747 - 763 ( 743 - 772)
   INTEGRAL    Likelihood =-12.52    Transmembrane  840 - 856 ( 835 - 856)
   INTEGRAL    Likelihood =-11.20    Transmembrane  447 - 463 ( 440 - 466)
   INTEGRAL    Likelihood = -5.79    Transmembrane  351 - 367 ( 346 - 372)
   INTEGRAL    Likelihood = -4.25    Transmembrane  517 - 533 ( 516 - 537)
   INTEGRAL    Likelihood = -1.49    Transmembrane  397 - 413 ( 396 - 413)
   INTEGRAL    Likelihood = -0.96    Transmembrane  799 - 815 ( 799 - 817)

----- Final Results -----
              bacterial membrane --- Certainty=0.6052(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9349> which encodes amino acid sequence <SEQ ID 9350> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAB89436 GB:AE000977 A. fulgidus predicted coding region AF1820
           [Archaeoglobus fulgidus]
 Identities = 100/483 (20%), Positives = 210/483 (42%), Gaps = 61/483 (12%)

Query: 351 LFPIILYLVAALVTLTTMTRFVEEERTNAGILKALGYSDRQVIFKFIIYGFIAGTLGTTL 410
           LFP    LV+  +T   ++R    +  N  +++ALG++  +++  ++ Y +  G   +T
Sbjct: 276 LFPAFFILVSIFMTYALLSRIFRLQLGNIAVMRALGFTRNEIMLHYLQYPLLMGFFASTA 335

Query: 411 GIIGGHYLLPRIISDIISKDLTIPNTQYHLFLNYSLLAFVFSLLSIVLPVFVI------- 463
           G++ G +    + S  I+  L +P       L    L+ +   L+  +  F++
Sbjct: 336 GLVAGFFASQLLTSQYIT-FLNLPYYVSKPHLEVYSLSLMAGTLTPTISGFLVAYQASRV 394

Query: 464 ----TRRELKEKAAFLLLPKPPAKGSKIALEYINWIWKKLSFTQKVTARNIFRYKQRMIM 519
                 R   E AA  +  +  A  S+I      W   ++    ++  RNIFR K+R  +
Sbjct: 395 DIVKALRGYAEVAAVSFIARIDALFSRI------W---RMRLIFRLALRNIFRSKRRTAI 445

Query: 520 TIFGVAGSVALLFSGLGIQSSLKQTVNEHFGRIMPYDILLTYNTNASPPKILELLSKDSK 579
```

```
           +IF +    +L+ + +    S    +   FG++  YDI ++           E+L K  K
Sbjct: 446 SIFSIVACTSLILNSMVFVDSFDYVMQLQFGKVYAYDIKVSLEGYDGK----EVLEKVRK 501

Query: 580 IDKY--------QPIHLENLDESIPGQINKQSISLFITDKKQLLPFIYLQEATTNKSLHL 631
           +D           PI++E   E++P        +L I    Q L  +Y  E        +
Sbjct: 502 MDGVLFAEPAVEMPIYVEKGGEAVP--------TLLIASNFQTLYNVYNAEG----EKLI 549

Query: 632 NNKGIIISKKLAQFYHVNTGDFIHL------SHSQTLPSRKLKITGVVNANVGHYIFMTK 685
            ++GII SK   +   +  G+ + +        ++      + +  V  A++
Sbjct: 550 PSEGIIFSKTAMKNLSLVEGEKVSVYTEFGKLEAEVEDVEMIPLLSVATASL-------- 601

Query: 686 QYYRTIFKKEAKDNAFLVKLTKHKIANNLAEKLLEINGVESLTQNALQLASVEAVVRSLD 745
            Y+   I   +   N  +V   + +IA  +AEK+ +++GV+ ++       S+E ++
Sbjct: 602 DYFSRISGVDG-FNRIVVDADEGRIA-EIAEKIRQMDGVKKVSTVIEAQESIEELMGFFY 659

Query: 746 GSMTILVVVSLLLAIVILYNLTNINLAERKRELSTIKVLGFYNEEVTLYIYRETIILSTI 805
             +   +   + L    ++N T+I++ ER REL+T+++LG+ + E+ + +  E + ++ +
Sbjct: 660 AFIAFSLFFGVSLGFAAVFNTTSISVIERSRELATLRMLGYTSREIIISLILENLFVAIL 719

Query: 806 GVI 808
           G++
Sbjct: 720 GLV 722
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1911> which encodes the amino acid sequence <SEQ ID 1912>. Analysis of this protein sequence reveals the following:

```
        Possible site: 34
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-14.33   Transmembrane  749 - 765 ( 739 - 775)
   INTEGRAL    Likelihood =-10.88   Transmembrane  845 - 861 ( 834 - 865)
   INTEGRAL    Likelihood = -6.64   Transmembrane  350 - 366 ( 344 - 369)
   INTEGRAL    Likelihood = -6.53   Transmembrane   22 -  38 (  19 -  42)
   INTEGRAL    Likelihood = -6.32   Transmembrane  520 - 536 ( 515 - 537)
   INTEGRAL    Likelihood = -4.99   Transmembrane  446 - 462 ( 445 - 465)
   INTEGRAL    Likelihood = -2.92   Transmembrane  396 - 412 ( 395 - 413)
   INTEGRAL    Likelihood = -0.80   Transmembrane  800 - 816 ( 800 - 819)

----- Final Results -----
               bacterial membrane --- Certainty=0.6731(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:AAB89436 GB:AE000977 A. fulgidus predicted coding region AF1820
           [Archaeoglobus fulgidus]
 Identities = 101/542 (18%), Positives = 237/542 (43%), Gaps = 42/542 (7%)

Query: 350 IFPVVLYLVAALVAFTTMTRYVDEERTSSGLLKAIGYSNKDISLKFLIYGLLASFLGTTL 409
           +FP    LV+  + +  ++R    +  +  +++A+G++  +I L +L Y LL  F  +T
Sbjct: 276 LFPAFFILVSIFMTYALLSRIFRLQLGNIAVMRALGFTRNEIMLHYLQYPLLMGFFASTA 335

Query: 410 GIIGGTYLLSTLISEILTGA---LTIGKTHLYSYWFYNGIAYLLAMLSAVLPAYLIVKKE 466
           G++ G +    L S+ +T       + K HL  Y        L   +S  L AY   + +
Sbjct: 336 GLVAGFFASQLLTSQYITFLNLPYYVSKPHLEVYSLSLMAGTLTPTISGFLVAYQASRVD 395

Query: 467 LFLN-------AAQLLLPKPPSKGAKIWLEHLTFVWKALSFTHKVTIRNIFRYKQRMLMT 519
           +          AA    + +  +  ++IW   L F         ++ +RNIFR K+R  ++
Sbjct: 396 IVKALRGYAEVAAVSFIARIDALFSRIWRMRLIF---------RLALRNIFRSKRRTAIS 446

Query: 520 IVGVAGSVALLFAGLGIQSSLAKVVEHQFGDLTTYDILAVGSAKATATEQTDLASYLKQE 579
           I  +    +L+   +    S   V++ QFG +  YDI            +  L  Y  +E
Sbjct: 447 IFSIVACTSLILNSMVFVDSFDYVMQLQFGKVYAYDI------------KVSLEGYDGKE 494

Query: 580 PITGYQKVSYASLTLPVKGLP---DKQSISILSSS-ATSLSPYFNLLDSQEQKKVPIPTS 635
            +   +K+       P   +P   +K   ++ +   A++    +N+ +++ +K   IP+
Sbjct: 495 VLEKVRKMDGVLFAEPAVEMPIYVEKGGEAVPTLLIASNFQTLYNVYNAEGEKL--IPSE 552
```

```
Query: 636 GVLISEKLASYYKVKPGDQLVLTDRKGQSYKVTIKQVIDMTVGHYLIMSDTYFKNHFKGL 695
           G++ S+       +  G+++ +    G+     ++  ++      L+   T   ++F  +
Sbjct: 553 GIIFSKTAMKNLSLVEGEKVSVYTEFGK-----LEAEVEDVEMIPLLSVATASLDYFSRI 607

Query: 696 EAAPAYLIKVKDKDSKHIKETASDLLTLKAIRAVSQNVNHIKSVQLVVTSLNQVMTLLVF 755
                +   V D D   I E A  +  +  ++ VS  +   +S++ ++      +   +F
Sbjct: 608 SGVDGFNRIVVDADEGRIAEIAEKIRQMDGVKKVSTVIEAQESIEELMGFFYAFIAFSLF 667

Query: 756 LSILLAIVILYNLTTINIAERIRELSTIKVLGFYDQEVTLYIYRETISLSLVGILLGIYL 815
             + L    ++N T+I++ ER REL+T+++LG+  +E+ + +  E + ++++G++  + +
Sbjct: 668 FGVSLGFAAVFNTTSISVIERSRELATLRMLGYTSREIIISLILENLFVAILGLVFALPI 727

Query: 816 GKGLHTYIMTMISTGDIQFGVKVDAYVYLVPILVILSLLAVLGIWVNRHLKKVDMLEALK 875
                 +  +   +      + +    +   +L + +++ +  +   R + ++D+ +  K
Sbjct: 728 AYSTAYFFFSSFESELYYMPMVIYPRTFAATVLAVFAIILLALLPSARRVSEMDIAKVTK 787

Query: 876 SI 877
            I
Sbjct: 788 EI 789
```

An alignment of the GAS and GBS proteins is shown below:

```
Identities = 377/857 (43%), Positives = 543/857 (62%), Gaps = 7/857 (0%)

Query: 3   KTFWKDIYRSITTSKGRFSSILLLMMLGSFAFIGLKVSAPNMQRTAQNYLAHHHVMDITV 62
           KT WKDI R+I  SKGRF S+  LM LGSFA +GLKV+ P+M+RTA  YL  H VMD+TV
Sbjct: 4   KTLWKDILRAIKNSKGRFISLFFLMALGSFALVGLKVTGPDMERTASRYLERHQVMDLTV 63

Query: 63  FNSWGLDKHDQTVLESLKGSQVEFSYFVDTTPQQNSKSYRLYSNTKTISTFDLVKGRLPL 122
             S    + D+  L++LKG+ +E+ + +D +   N KS RLYS  K +S   LVKG  P
Sbjct: 64  LASHQFSQADKQELDTLKGAHLEYGHLLDVSLTSNQKSLRLYSVPKKVSKPVLVKGSWPK 123

Query: 123 NKSEIALSFQERKKYAIGDKINFKQDKNKLFSNTGPLTIVGFVNSTEIWSKTNLGSSQTG 182
            ++++ LS    K Y IGD++        L + T    +VGF NS+E+WSK+NLGSS TG
Sbjct: 124 RETDLVLSSSLAKNYQIGDELAVTSPMEGLLTTTH-FQVVGFANSSEVWSKSNLGSSSTG 182

Query: 183 DGDLDSYGVLDKTAFHSPVYTMARVTFKDLRLINPFSISYKEKVAKYQEKVSRKLNIHNK 242
           DG L +Y  ++   F S  + + R+ F  LRL N FS  Y+++V + Q  +   L  + +
Sbjct: 183 DGSLYAYAFVNPNVFKS-AFNLLRIRFSHLRLTNAFSKDYQKRVTQNQAHLDNLLKDNGQ 241

Query: 243 IRYTKTKKESLRKIDEEEKSLLKAQKQINRLDNDSLAMPLSQRQAIQMKIKQDRLSLLKR 302
            RY   + +    +     +L K   ++   + +   +  S  Q  + +I+Q + +L K
Sbjct: 242 KRYDDLQNQYDLALKNGRAALAKETVKLAASEENLTFLEGSALQEAKHQIEQGKQALAKE 301

Query: 303 TKELLKLRHNTQIMESPQIIVYNRTTFPGGQGYNTFDSSTNSTSKISNLFPIILYLVAAL 362
            K+L +++     +E P  + YNR+T PGG+GY+T+ +ST S S + N+FP++LYLVAAL
Sbjct: 302 EKQLEQVQATKDKLEKPSYLTYNRSTLPGGEGYHTYATSTTSISNVGNIFPVVLYLVAAL 361

Query: 363 VTLTTMTRFVEEERTNAGILKALGYSDRQVIFKFIIYGFIAGTLGTTLGIIGGHYLLPRI 422
           V  TTMTR+V+EERT++G+LKA+GYS++ +  KF+IYG +A  LGTTLGIIGG YLL  +
Sbjct: 362 VAFTTMTRYVDEERTSSGLLKAIGYSNKDISLKFLIYGLLASFLGTTLGIIGGTYLLSTL 421

Query: 423 ISDIISKDLTIPNTQYHLFLNYSLLAFVFSLLSIVLPVFVITRRELKEKAAFLLLPKPPA 482
           IS+I++  LTI  T  + +  Y+ +A++ ++LS VLP ++I ++EL   AA LLLPKPP+
Sbjct: 422 ISEILTGALTIGKTHLYSYWFYNGIAYLLAMLSAVLPAYLIVKKELFLNAAQLLLPKPPS 481

Query: 483 KGSKIALEYINWIWKKLSFTQKVTARNIFRYKQRMIMTIFGVAGSVALLFSGLGIQSSLK 542
           KG+KI LE++ ++WK LSFT KVT RNIFRYKQRM+MTI GVAGSVALLF+GLGIQSSL
Sbjct: 482 KGAKIWLEHLTFVWKALSFTHKVTIRNIFRYKQRMLMTIVGVAGSVALLFAGLGIQSSLA 541

Query: 543 QTVNEHFGRIMPYDILLTYNTNASPPKILELLS--KDSKIDKYQPIHLENLDESIPGQIN 600
           + V   FG +  YDIL   +  A+  +  +L S  K   I  YQ +   +L   + G  +
Sbjct: 542 KVVEHQFGDLTTYDILAVGSAKATATEQTDLASYLKQEPITGYQKVSYASLTLPVKGLPD 601

Query: 601 KQSISLFITDKKQLLPFIYLQEATTNKSLHLNNKGIIISKKLAQFYHVNTGDFIHLSHSQ 660
           KQSIS+  +    L P+  L ++   K + +   G++IS+KLA +Y V  GD + L+  +
Sbjct: 602 KQSISILSSSATSLSPYFNLLDSQEQKKVPIPTSGVLISEKLASYYKVKPGDQLVLTDRK 661
```

```
Query: 661 TLPSRKLKITGVVNANVGHYIFMTKQYYRTIFKKEAKDNAFLVKL--TKHKIANNLAEKL 718
              S K+ I  V++  VGHY+ M+  Y++  FK      A+L+K+     K     A  L
Sbjct: 662 G-QSYKVTIKQVIDMTVGHYLIMSDTYFKNHFKGLEAAPAYLIKVKDKDSKHIKETASDL 720

Query: 719 LEINGVESLTQNALQLASVEAVVRSLDGSMTILVVVSLLLAIVILYNLTNINLAERKREL 778
           L +  + +++QN   + SV+ VV SL+  MT+LV +S+LLAIVILYNLT IN+AER REL
Sbjct: 721 LTLKAIRAVSQNVNHIKSVQLVVTSLNQVMTLLVFLSILLAIVILYNLTTINIAERIREL 780

Query: 779 STIKVLGFYNEEVTLYIYRETIILSTIGVILGTISGTYLHRQMMLLIGSDQILFGEKVSP 838
           STIKVLGFY++EVTLYIYRETI LS +G++LG   G  LH  +M +I +  I FG KV
Sbjct: 781 STIKVLGFYDQEVTLYIYRETISLSLVGILLGIYLGKGLHTYIMTMISTGDIQFGVKVDA 840

Query: 839 TTFIIPISVVVIILXXL 855
             +++PI V++ +L  L
Sbjct: 841 YVYLVPILVILSLLAVL 857
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 616

A DNA sequence (GBSx0656) was identified in *S.agalactiae* <SEQ ID 1913> which encodes the amino acid sequence <SEQ ID 1914>. This protein is predicted to be ABC transporter, ATP-binding protein. Analysis of this protein sequence reveals the following:

```
Possible site: 60
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2757(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAB89431 GB:AE000977 ABC transporter, ATP-binding protein
           [Archaeoglobus fulgidus]
 Identities = 112/230 (48%), Positives = 167/230 (71%)

Query: 4   IEMKHSYKRYQTGETEIVANNDISFSIERGELVVILGASGAGKSTVLNILGGMDSNSEGE 63
           + ++  +K YQ G+ E+ A   I+  IERGE +V+LG SG GK+T+LNI+GG+D  + G
Sbjct: 2   LRLEDVWKVYQMGKVEVSALRGINLEIERGEFMVVLGPSGCGKTTMLNIIGGIDRPTRGR 61

Query: 64  VLIDGKNIANYTIRELTRYRRYDVGFVFQFYNLVPNLTALENVELASEIVPKALDAQQAL 123
           V+ DGK+I NY     LT +RR +VGF+FQF+NL+P LTA ENVE+A+++V    D + L
Sbjct: 62  VIFDGKDITNYNEDRLTMHRRNNVGFIFQFFNLIPTLTARENVEIAADLVESPRDVDEVL 121

Query: 124 ENVGLGHRINHFPAQLSGGEQQRVAIARAIAKKPKLLLCDEPTGALDYQTGKQVLAILQK 183
           + VGL  R  HFPA+LSGGEQQRVAIARA+ K P ++L DEPTG+LD++TGK VL ++++
Sbjct: 122 KMVGLADRAEHFPAELSGGEQQRVAIARALVKNPPIILADEPTGSLDFETGKAVLKVMRE 181

Query: 184 MAQSKETTVIIVTHNTALAPIANRVIHMHDSKISDIVINENPSDIQNIEY 233
           + + +  T ++VTHN+A+A IA+RV+++ D K+  +  N +P+D   I++
Sbjct: 182 INRKEGITFVLVTHNSAIAAIADRVVYLRDGKVERVERNLHPADPDEIQW 231
```

There is also homology to *SEQ ID 1354*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 617

A DNA sequence (GBSx0657) was identified in *S.agalactiae* <SEQ ID 1915> which encodes the amino acid sequence <SEQ ID 1916>. This protein is predicted to be DNA topoisomerase I (topA). Analysis of this protein sequence reveals the following:

```
Possible site: 34
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4716(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9821> which encodes amino acid sequence <SEQ ID 9822> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB13485 GB:Z99112 DNA topoisomerase I [Bacillus subtilis]
 Identities = 442/690 (64%), Positives = 535/690 (77%), Gaps = 10/690 (1%)

Query: 27  LVIVESPAKAKTIEKYLGRNYKVVASVGHIRDLKKSSMSIDFENNYEPQYINIRGKGPLI 86
            LVIVESPAKAKTIE+YLG+ YKV AS+GH+RDL KS M +D E N+EP+YI IRGKGP++
Sbjct: 5   LVIVESPAKAKTIERYLGKKYKVKASMGHVRDLPKSQMGVDIEQNFEPKYITIRGKGPVL 64

Query: 87  NDLKKEAKKAKKVYLASDPDREGEAISWHLAHILDLDKEDRNRVVFNEITKDAVKNAFVE 146
             +LK  AKKAKKVYLA+DPDREGEAI+WHLAH LDLD     RVVFNEITKDA+K +F
Sbjct: 65  KELKTAAKKAKKVYLAADPDREGEAIAWHLAHSLDLDLNSDCRVVFNEITKDAIKESFKH 124

Query: 147 PRQINMDLVDAQQARRVLDRIVGYSISPILWKKVKKGLSAGRVQSVALKLIIDRENEIKA 206
            PR INMDLVDAQQARR+LDR+VGY ISPILWKKVKKGLSAGRVQSVAL+LIIDRE EI
Sbjct: 125 PRMINMDLVDAQQARRILDRLVGYKISPILWKKVKKGLSAGRVQSVALRLIIDREKEIND 184

Query: 207 FQPEEYWTIDGSFKKGTRKFNATFYGLDGKKFKLSNNEDVKTVLKRIKTDEFLVEKVEKK 266
            F+PEEYWTIDG+F KG   F A+F+G +GKK  L++  DVK +L ++K +++ VEKV KK
Sbjct: 185 FKPEEYWTIDGTFLKGQETFEASFFGKNGKKLPLNSEADVKEILSQLKGNQYTVEKVTKK 244

Query: 267 ERRRNAPLPYTTSSLQQDAANKINFRTRKTMMIAQQLYEGLSLGTAGHQGLITYMRTDST 326
            ER+RN  LP+TTS+LQQ+AA K+NFR +KTMMIAQQLYEG+ LG  G  GLITYMRTDST
Sbjct: 245 ERKRNPALPFTTSTLQQEAARKLNFRAKKTMMIAQQLYEGIDLGREGTVGLITYMRTDST 304

Query: 327 RISPLAQNEATEFITNRFGANYSKHGNK-VKNASGAQDAHEAIRPSSVNHTPESIAKYLD 385
            RIS  A +EA  FI   +G  +     K  K    AQDAHEAIRP+SV   P  +   L
Sbjct: 305 RISNTAVDEAAAFIDQTYGKEFLGGKRKPAKKNENAQDAHEAIRPTSVLRKPSELKAVLG 364

Query: 386 KDQLKLYTLIWNRFIASQMTAAVFDTMKVNLTQNGVTFIANGSQVKFDGYMAVYND---- 441
            +DQ++LY LIW RF+ASQM  AV DTM V+LT NG+TF ANGS+VKF G+M VY +
Sbjct: 365 RDQMRLYKLIWERFVASQMAPAVLDTMSVDLTNNGLTFRANGSKVKFSGFMKVYVEGKDD 424

Query: 442 --TDKNKMLPDMEEGESVKKVNTNPEQHFTQPPARFSEASLIKTLEENGVGRPSTYAPTL 499
               +K++MLPD++EG++V   +  PEQHFTQPP R++EA L+KTLEE G+GRPSTYAPTL
Sbjct: 425 QMEEKDRMLPDLQEGDTVLSKDIEPEQHFTQPPPRYTEARLVKTLEERGIGRPSTYAPTL 484

Query: 500 ETIQKRYYVKLAAKRFEPTELGEIVNSLIVEFFPDIVDVTFTAEMEGKLDEVEIGKEQWQ 559
            +TIQ+R YV L  KRF PTELG+IV  LI+EFFP+I++V FTA+ME  LD VE G  +W
Sbjct: 485 DTIQRRGYVALDNKRFVPTELGQIVLDLIMEFFPEIINVEFTAKMERDLDHVEEGNTEWV 544

Query: 560 KIIDEFYKPFEKELAKAETEMEKIQIKDEPAGFDCELCGSPMVIKLGRYGKFYACSNFPE 619
            KIID FY  FEK + KAE+EM++++I+ E AG DCELC SPMV K+GRYGKF ACSNFP+
Sbjct: 545 KIIDNFYTDFEKRVKKAESEMKEVEIEPEYAGEDCELCSSPMVYKMGRYGKFLACSNFPD 604

Query: 620 CHNTKAITKEIGVICPICQKGQVIERKTKRNRIFYGCDRYPECEFTSWDKPIGRTCPKSN 679
            C NTK I K+IGV CP C +G ++ERK+K+ R+FYGCDRYP+CEF SWDKPI R CPK
Sbjct: 605 CRNTKPIVKQIGVKCPSCGEGNIVERKSKKKRVFYGCDRYPDCEFVSWDKPIERKCPKCG 664
```

```
Query: 680 DFLVEKKVRGGGKQVVCSNEKCDYQEEKIK 709
             LVEKK++  G QV C   +CDY+EE  K
Sbjct: 665 KMLVEKKLK-KGIQVQC--VECDYKEEPQK 691
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1917> which encodes the amino acid sequence <SEQ ID 1918>. Analysis of this protein sequence reveals the following:

```
Possible site: 35
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.5445(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 595/704 (84%), Positives = 656/704 (92%), Gaps = 1/704 (0%)

Query: 6    TTTKTSTKKTSKKKSATAKKNLVIVESPAKAKTIEKYLGRNYKVVASVGHIRDLKKSSMS 65
            T  KT TKK++ KK +TAKKNLVIVESPAKAKTIEKYLGR+YKVVASVGHIRDLKKSSMS
Sbjct: 7    TKPKTGTKKSTTKKKSTAKKNLVIVESPAKAKTIEKYLGRSYKVVASVGHIRDLKKSSMS 66

Query: 66   IDFENNYEPQYINIRGKGPLINDLKKEAKKAKKVYLASDPDREGEAISWHLAHILDLDKE 125
            IDF+NNYEPQYINIRGKGPLIN LKKEAK AKKVYLASDPDREGEAISWHL+HIL LD +
Sbjct: 67   IDFDNNYEPQYINIRGKGPLINSLKKEAKAAKKVYLASDPDREGEAISWHLSHILGLDPQ 126

Query: 126  DRNRVVFNEITKDAVKNAFVEPRQINMDLVDAQQARRVLDRIVGYSISPILWKKVKKGLS 185
            D NRVVFNEITKDAVK+AFVEPRQI+MDLVD+QQARRVLDRIVGYSISPILWKKVKKGLS
Sbjct: 127  DNNRVVFNEITKDAVKHAFVEPRQIDMDLVDSQQARRVLDRIVGYSISPILWKKVKKGLS 186

Query: 186  AGRVQSVALKLIIDRENEIKAFQPEEYWTIDGSFKKGTRKFNATFYGLDGKKFKLSNNED 245
            AGRVQSVALKLIIDREN+IKAF P+EYW+IDG FKKGT+KF ATFYG++GKK KL NN D
Sbjct: 187  AGRVQSVALKLIIDRENDIKAFVPKEYWSIDGLFKKGTKKFQATFYGINGKKTKLDNNND 246

Query: 246  VKTVLKRIKTDEFLVEKVEKKERRRNAPLPYTTSSLQQDAANKINFRTRKTMMIAQQLYE 305
            VK VL ++  ++FLV KV+KKERRRNAPLPYTTSSLQQDAANKINFRTRKTMM+AQQLYE
Sbjct: 247  VKEVLAKLTNEDFLVSKVDKKERRRNAPLPYTTSSLQQDAANKINFRTRKTMMVAQQLYE 306

Query: 306  GLSLGTAGHQGLITYMRTDSTRISPLAQNEATEFITNRFGANYSKHGNKVKNASGAQDAH 365
            G+ LG  G QGLITYMRTDSTRISP+AQN+A +FI NRFGANYSKHGN+VKN SG QDAH
Sbjct: 307  GIHLGENGTQGLITYMRTDSTRISPVAQNDAAQFIINRFGANYSKHGNRVKNTSGVQDAH 366

Query: 366  EAIRPSSVNHTPESIAKYLDKDQLKLYTLIWNRFIASQMTAAVFDTMKVNLTQNGVTFIA 425
            EAIRPSSVNHTP+SIAKYL+KDQLKLYTLIWNRF+ASQMTAAVFDT+KVNL QNGV F+A
Sbjct: 367  EAIRPSSVNHTPDSIAKYLNKDQLKLYTLIWNRFVASQMTAAVFDTVKVNLEQNGVIFVA 426

Query: 426  NGSQVKFDGYMAVYNDTDKNKMLPDMEEGESVKKVNTNPEQHFTQPPARFSEASLIKTLE 485
            NGSQ+KFDGYMAVYND+DKNKMLP+M EGE+VKK++T+PEQHFTQPPAR+SEA+LIKTLE
Sbjct: 427  NGSQMKFDGYMAVYNDSDKNKMLPEMAEGETVKKISTSPEQHFTQPPARYSEATLIKTLE 486

Query: 486  ENGVGRPSTYAPTLETIQKRYYVKLAAKRFEPTELGEIVNSLIVEFFPDIVDVTFTAEME 545
            ENGVGRPSTYAPTLE IQ+RYYVKL+AKRFEPTELGEIVN LIVEFFPDIVDV FTAEME
Sbjct: 487  ENGVGRPSTYAPTLEVIQRRYYVKLSAKRFEPTELGEIVNKLIVEFFPDIVDVAFTAEME 546

Query: 546  GKLDEVEIGKEQWQKIIDEFYKPFEKELAKAETEMEKIQIKDEPAGFDCELCGSPMVIKL 605
            GKLD+VEIG+EQWQ +ID+FY+PF KEL KAE+E+EKIQIKDEPAGFDC++CG PMVIKL
Sbjct: 547  GKLDQVEIGEEQWQHVIDQFYQPFVKELNKAESEIEKIQIKDEPAGFDCDVCGHPMVIKL 606

Query: 606  GRYGKFYACSNFPECHNTKAITKEIGVICPICQKGQVIERKTKRNRIFYGCDRYPECEFT 665
            GR+GKFYACSNFPEC NTKAITKEIGV CP+C KGQVIERKTK+NRIFYGCD+YP+CEF
Sbjct: 607  GRFGKFYACSNFPECRNTKAITKEIGVTCPVCHKGQVIERKTKKNRIFYGCDQYPDCEFI 666

Query: 666  SWDKPIGRTCPKSNDFLVEKKVRGGGKQVVCSNEKCDYQEEKIK 709
            SWD PIGR CPKS D+L+EKKVR GGKQV+CSNE CDY+EEKIK
Sbjct: 667  SWDLPIGRACPKSGDYLIEKKVR-GGKQVMCSNETCDYKEEKIK 709
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 618

A DNA sequence (GBSx0658) was identified in *S.agalactiae* <SEQ ID 1919> which encodes the amino acid sequence <SEQ ID 1920>. Analysis of this protein sequence reveals the following:

```
Possible site: 43
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2578(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAD35341 GB:AE001708 DNA processing chain A [Thermotoga maritima]
 Identities = 97/231 (41%), Positives = 149/231 (63%), Gaps = 2/231 (0%)

Query: 51  FIENYKQLDLKKLRQEFKKFPV--LSILDSNYPLELKEIYNPPVLLFYQGNIELLSKPKL 108
            F+E   + +L++ ++  +K  V  +S  + +YP  L+EI  PP +LF +G+ ELL +  +
Sbjct: 41  FLEKCGKEELERQKELIRKHNVKLVSFWEDDYPQHLREIRYPPAVLFVRGDAELLKEKCV 100

Query: 109 AVVGARQASQIGCQSVKKIIKETNNQFVIVSGLARGIDTAAHVSALKNGGSSIAVIGSGL 168
             VVG R+ +  G    K+ +K  +  FVIVSG+A GID+ AH  AL +GG ++AV+G+G+
Sbjct: 101 GVVGTRRPTSYGVNVTKRFVKLLSEYFVIVSGMAFGIDSVAHKEALSSGGKTVAVLGTGV 160

Query: 169 DVYYPTENKKLQEYMSYNHLVLSEYFTGEQPLKFHFPERNRIIAGLCQGIVVAEAKMRSG 228
            DV YP  N++L   +  N  V+SEY  G +  K HFP RNRIIAGL   I+V EA ++SG
Sbjct: 161 DVVYPRSNERLFHEIVKNGCVVSEYPMGTRARKHHFPARNRIIAGLSDAIIVTEAPIKSG 220

Query: 229 SLITCERALEEGREVFAIPGNIIDGKSDGCHHLIQEGAKCIISGKDILSEY 279
            +LIT + ALE GR+VFA+PG+I    S+G ++LI+ GA  +   +D+ + +
Sbjct: 221 ALITVKFALESGRDVFAVPGDIDRKTSEGTNYLIKSGAYPLTDEEDLETHF 271
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1921> which encodes the amino acid sequence <SEQ ID 1922>. Analysis of this protein sequence reveals the following:

```
Possible site: 50
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2856(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 185/279 (66%), Positives = 238/279 (84%), Gaps = 1/279 (0%)

Query: 1   MNHFELFKLKKAGLTNLNIHNIINYLKKNSLTSLSVRNMAVVSKCKNPTFFIENYKQLDL 60
            +NHFEL+KLKKAGLTN NI NI++Y +K+   SLS+R+MAVVS CK+P+ FIE YKQLD+
Sbjct: 1   VNHFELYKLKKAGLTNKNILNILDY-QKHQEKSLSLRDMAVVSGCKHPSHFIEAYKQLDI 59

Query: 61  KKLRQEFKKFPVLSILDSNYPLELKEIYNPPVLLFYQGNIELLSKPKLAVVGARQASQIG 120
            + L+ EFK+FP +SILD +YP+ LKEIYNPPVLLF+QGN++LL KPKLA+VG+R++S  G
Sbjct: 60  QNLKMEFKQFPSISILDKHYPMALKEIYNPPVLLFFQGNLDLLEKPKLAIVGSRRSSDTG 119

Query: 121 CQSVKKIIKETNNQFVIVSGLARGIDTAAHVSALKNGGSSIAVIGSGLDVYYPTENKKLQ 180
             +SV+KI+KE  N+FVIVSGLARGIDT+AH++ LKNGG +IA+IG+GLD +YP EN++LQ
Sbjct: 120 VKSVRKILKELGNRFVIVSGLARGIDTSAHLACLKNGGQTIAIIGTGLDRFYPKENRELQ 179

Query: 181 EYMSYNHLVLSEYFTGEQPLKFHFPERNRIIAGLCQGIVVAEAKMRSGSLITCERALEEG 240
```

```
            ++  NHLVL+EY  GE+ L +HFPERNRIIAGL +GI+V EAK RSGSLITC+  +EEG
Sbjct: 180 TFLGKNHLVLTEYGPGEEALSYHFPERNRIIAGLSRGILVVEAKNRSGSLITCQIGIEEG 239

Query: 241 REVFAIPGNIIDGKSDGCHHLIQEGAKCIISGKDILSEY 279
           R++FA+PGNI+DGKS+GC  LI+EGA C+ SG DILSEY
Sbjct: 240 RDIFAVPGNILDGKSEGCLQLIKEGATCVTSGMDILSEY 278
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 619

A DNA sequence (GBSx0659) was identified in *S.agalactiae* <SEQ ID 1923> which encodes the amino acid sequence <SEQ ID 1924>. This protein is predicted to be lipoprotein (ceuE). Analysis of this protein sequence reveals the following:

```
Possible site: 24
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
                bacterial outside --- Certainty=0.3000(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAA06500 GB:AJ005352 lipoprotein [Staphylococcus aureus]
 Identities = 122/348 (35%), Positives = 201/348 (57%), Gaps = 16/348 (4%)

Query: 1   MTKKLIIAILALCTILTTSQAVLAKEKSQ--------TVTIKNNYSVYIKKEKRDKPDNK 52
           M K ++  +LA+  +L        KE+S+        TV I+NNY +  + EK+D  D K
Sbjct: 1   MKKTVLYLVLAVMFLLAACGNNSDKEQSKSETKGSKDTVKIENNYKM--RGEKKDGSDAK 58

Query: 53  KQISETLKVPLKPKKVVVFDMGALDTITALGAEKSVIGIPKAKNALSLLPNNVKSVYKAK 112
           K + ET++VP  P+  VV D GALD +  +G    V  +PK +   SL PN ++S +K
Sbjct: 59  K-VKETVEVPKNPENAVVLDYGALDVMKEMGLSDKVKALPKGEGGKSL-PNFLES-FKDD 115

Query: 113 RYQDVGSLFEPNFEAIARMQPDVVFLGARMASVDNIEKLKEAAPKAALVYAGVDSKKVFD 172
           +Y +VG+L E NF+ IA  +P+V+F+  R A+  N+++ K+AAPKA +VY G D K +
Sbjct: 116 KYTNVGNLKEVNFDKIAATKPEVIFISGRTANQKNLDEFKKAAPKAKIVYVGADEKNLIG 175

Query: 173 KGVAERVTMLGKIFDQNKKAKTFNKDIAQAVLKLQKTIEKKGKPTALFVMANSGELLTQS 232
             + +    +GKI+D+  KAK  NKD+   +  ++   +   K T ++++ N GEL T
Sbjct: 176 S-MKQNTENIGKIYDKEVKAKELNKDLDNKIASMKDKTKNFNK-TVMYLLVNEGELSTFG 233

Query: 233 PSGRFGW-IFSVGGFKAVNENEKLSSHGTPVSYEYIAEKNPNYLFVLDRGATIGQGASSK 291
           P GRFG  ++   GF AV++    S+HG  VS EY+ ++NP+ +  +DRG  +   +++K
Sbjct: 234 PKGRFGGLVYDTLGFNAVDKKVSNSNHGQNVSNEYVNKENPDVILAMDRGQAVSGKSTAK 293

Query: 292 ELFNNDVIKATDAVKNKRVHEVDGKDWYINSGGSRVTLRMIKDVQNFV 339
           +  NN V+K   A+K  +V+ +D K WY  +G +  T++ I+++   V
Sbjct: 294 QALNNPVLKNVKAIKEDKVYNLDPKLWYFAAGSTTTTIKQIEELDKVV 341
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1925> which encodes the amino acid sequence <SEQ ID 1926>. Analysis of this protein sequence reveals the following:

```
Possible site: 32
>>> May be a lipoprotein

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 57/255 (22%), Positives = 104/255 (40%), Gaps = 30/255 (11%)

Query: 66  KKVVVFDMGALDTITALGAEKSVIGIPKAKNALSLLPNNVKSVYKAKRYQDVGSLFEPNF 125
            +++V   +  +D    L  +  ++G+  +K  L  LP    +V +      VG    P+
Sbjct: 45  QRIVATSVAVVDICDRLNLD--LVGVCDSK--LYTLPKRYDAVKR------VGLPMNPDI 94

Query: 126 EAIARMQPDVVFLGARMASVDNIEKLKEAAPKAALVYAGVDSKKVFDKGVAERVTMLGKI 185
            E IA ++P  +     +      E L+    K    Y  ++ + V  +G+ + +  LG +
Sbjct: 95  ELIASLKPTWILSPNSLQ-----EDLEPKYQKLDTEYGFLNLRSV--EGMYQSIDDLGNL 147

Query: 186 FDQNKKAKTFNKDIAQAVLKLQKTIEKKGKPTALFVMANSGELLTQSPSGRFGWIFSVGG 245
            F + ++AK   +         Q   + K KP  L +M   G  L  +     G +  + G
Sbjct: 148 FQRQQEAKELRQQYQDYYRAFQAKRKGKKKPKVLILMGLPGSYLVATNQSYVGNLLDLAG 207

Query: 246 FKAV---NENEKLSSHGTPVSYEYIAEKNPNYLFVLDRGATIGQGAS---SKELFNNDVI 299
             + V   +E E LS++      E +  K P+   +L     I         KE   ND+
Sbjct: 208 GENVYQSDEKEFLSANP-----EDMLAKEPD--LILRTAHAIPDKVKVMFDKEFAENDIW 260

Query: 300 KATDAVKNKRVHEVD 314
            K   AVK  +V+++D
Sbjct: 261 KHFTAVKEGKVYDLD 275
```

SEQ ID 1924 (GBS181) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 39 (lane 5; MW 38.7kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 46 (lane 3; MW 64kDa).

The GBS181-GST fusion product was purified (Figure 204, lane 9) and used to immunise mice. The resulting antiserum was used for FACS (Figure 299), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 620

A DNA sequence (GBSx0660) was identified in *S.agalactiae* <SEQ ID 1927> which encodes the amino acid sequence <SEQ ID 1928>. This protein is predicted to be iron(III) ABC transporter, ATP-binding protein. Analysis of this protein sequence reveals the following:

```
Possible site: 40
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3231(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB12190 GB:Z99106 similar to ferrichrome ABC transporter
          (ATP-binding protein) [Bacillus subtilis]
 Identities = 125/247 (50%), Positives = 187/247 (75%)

Query: 1   MIQINNLHKFYGQKEILKDINISIPKGKVTAILGPNGSGKSTLLSCISRLEPYDNGEIFL 60
            M+++ N+ K YG K +L++ +++I KGK+T+ +GPNG+GKSTLLS +SRL   D+GEI++
Sbjct: 1   MVEVRNVSKQYGGKVVLEETSVTIQKGKITSFIGPNGAGKSTLLSIMSRLIKKDSGEIYI 60

Query: 61  DKVPLAHYSSNDLAKTLAILRQSNHLTLKIKVRDLIGFGRFPYSKGRLSQKDKAVIESVI 120
            D   +    S +LAK ++IL+Q+N + +++ ++DL+ FGRFPYS+GRL+++D   I   +
Sbjct: 61  DGQEIGACDSKELAKKMSILKQANQINIRLTIKDLVSFGRFPYSQGRLTEEDWVHINQAL 120
```

```
Query: 121 SYMDLNDIADEFINNLSGGQIQRAFIAMTMAQDTQYICLDEPLNNLDMKYAVQMMDLIKR 180
           SYM L DI D++++ LSGGQ QRAFIAM +AQDT YI LDEPLNNLDMK++V++M L+KR
Sbjct: 121 SYMKLEDIQDKYLDQLSGGQCQRAFIAMVIAQDTDYIFLDEPLNNLDMKHSVEIMKLLKR 180

Query: 181 YAYEFNKTIVIIIHDINFATHYADNVVALKEGQVVTCGTVEDVMQEKILSHLFDMPIRIE 240
              E  KTIVI+IHDINFA+ Y+D +VALK G++V  G  E++++  +L  ++DM I I+
Sbjct: 181 LVEELGKTIVIVIHDINFASVYSDYIVALKNGRIVKEGPPEEMIETSVLEEIYDMTIPIQ 240

Query: 241 TVDGKPI 247
           T+D + I
Sbjct: 241 TIDNQRI 247
```

There is also homology to SEQ ID 1930.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 621

A DNA sequence (GBSx0661) was identified in *S.agalactiae* <SEQ ID 1931> which encodes the amino acid sequence <SEQ ID 1932>. Analysis of this protein sequence reveals the following:

```
Possible site: 33
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood =-12.74    Transmembrane  271 - 287 ( 266 - 295)
   INTEGRAL    Likelihood = -8.55    Transmembrane   49 -  65 (  47 -  69)
   INTEGRAL    Likelihood = -8.07    Transmembrane  185 - 201 ( 178 - 207)
   INTEGRAL    Likelihood = -7.70    Transmembrane  112 - 128 ( 105 - 132)
   INTEGRAL    Likelihood = -7.38    Transmembrane  231 - 247 ( 227 - 261)
   INTEGRAL    Likelihood = -2.50    Transmembrane  139 - 155 ( 135 - 156)
   INTEGRAL    Likelihood = -1.97    Transmembrane  302 - 318 ( 301 - 319)

----- Final Results -----
               bacterial membrane --- Certainty=0.6095(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB12189 GB:Z99106 similar to ferrichrome ABC transporter
           (permease) [Bacillus subtilis]
 Identities = 138/315 (43%), Positives = 222/315 (69%), Gaps = 6/315 (1%)

Query: 9   KLLILLILLIAAIILFLIYGIPTDANEFLIIYILKTRYQKLIALILVGICIGSSSLIFQT 68
           K+ +L+ L I  I LFL Y +          Y L  R +K+ A++L G  I  S++IFQT
Sbjct: 6   KIALLVGLAIVCIGLFLFYDLGNWD------YTLPRRIKKVAAIVLTGGAIAFSTMIFQT 59

Query: 69  LTNNRLLTPSIIGLDSLYILIQTGLMYLIGAQRVIKFSSFSSFLLSLLLMVGFAYLLFTI 128
           +TNNR+LTPSI+GLDSLY+LIQTG+++L G+  ++  +   +F++S+LLM+ F+ +L+ I
Sbjct: 60  ITNNRILTPSILGLDSLYMLIQTGIIFLFGSANMVIMNKNINFIISVLLMILFSLVLYQI 119

Query: 129 LFRNKKQSLYFVLLAGLIFNTLFSSISSFIQAIMDPNDFMILQNQLFASFNAINTKILWI 188
           +F+ + ++++F+LL G++F TLFSS+SSF+Q ++DPN+F ++Q+++FASFN INT +LW+
Sbjct: 120 MFKGEGRNIFFLLLIGIVFGTLFSSLSSFMQMLIDPNEFQVVQDKMFASFNNINTDLLWL 179

Query: 189 SFIIIVVSFVINWPFIKELDVLLLGKENAISLGISYQKLTTRFFLWLALMVAIATALVGP 248
           +FII +++ V  W F K  DVL LG+E+A++LGI Y K+   +   +A++V+++TALVGP
Sbjct: 180 AFIIFLLTGVYVWRFTKFFDVLSLGREHAVNLGIDYDKVVKQMLIVVAILVSVSTALVGP 239

Query: 249 ITFLGLLVAHITYHSFHTFRHQILVPIAIVICIFTLVLGQHLVQNLLHLTVQLSVLLNLI 308
           I FLGLLV ++      T++H  L+  ++ I I  LV GQ +V+ +   +  LSV++N
Sbjct: 240 IMFLGLLVVNLAREFLKTYKHSYLIAGSVFISIIALVGGQFVVEKVFTFSTTLSVIINFA 299

Query: 309 GGSYFIFTLIKGRKN 323
           GG YFI+ L+K  K+
```

```
Sbjct: 300 GGIYFIYLLLKENKS 314
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1933> which encodes the amino acid sequence <SEQ ID 1934>. Analysis of this protein sequence reveals the following:

```
Possible site: 58
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-13.64   Transmembrane   33 -  49 (  26 -  61)
   INTEGRAL    Likelihood = -8.97   Transmembrane  259 - 275 ( 246 - 286)
   INTEGRAL    Likelihood = -8.65   Transmembrane  296 - 312 ( 294 - 316)
   INTEGRAL    Likelihood = -8.39   Transmembrane   83 -  99 (  78 - 104)
   INTEGRAL    Likelihood = -6.26   Transmembrane  212 - 228 ( 210 - 231)
   INTEGRAL    Likelihood = -4.04   Transmembrane  113 - 129 ( 110 - 132)
   INTEGRAL    Likelihood = -3.61   Transmembrane  140 - 156 ( 134 - 157)
   INTEGRAL    Likelihood = -2.71   Transmembrane  165 - 181 ( 165 - 181)
   INTEGRAL    Likelihood = -1.06   Transmembrane  327 - 343 ( 327 - 343)
   INTEGRAL    Likelihood = -0.22   Transmembrane   50 -  66 (  50 -  66)

----- Final Results -----
              bacterial membrane --- Certainty=0.6456(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related sequence was also identified in GAS <SEQ ID 9175> which encodes the amino acid sequence <SEQ ID 9176>. Analysis of this protein sequence reveals the following:

```
     Possible site: 49
>>> Seems to have no N-terminal signal sequence
     INTEGRAL    Likelihood =-13.64   Transmembrane   24 -  40 (  17 -  52)
     INTEGRAL    Likelihood = -8.97   Transmembrane  250 - 266 ( 237 - 277)
     INTEGRAL    Likelihood = -8.65   Transmembrane  287 - 303 ( 285 - 307)
     INTEGRAL    Likelihood = -8.39   Transmembrane   74 -  90 (  69 -  95)
     INTEGRAL    Likelihood = -6.26   Transmembrane  203 - 219 ( 201 - 222)
     INTEGRAL    Likelihood = -4.04   Transmembrane  104 - 120 ( 101 - 123)
     INTEGRAL    Likelihood = -3.61   Transmembrane  131 - 147 ( 125 - 148)
     INTEGRAL    Likelihood = -2.71   Transmembrane  156 - 172 ( 156 - 172)
     INTEGRAL    Likelihood = -1.06   Transmembrane  318 - 334 ( 318 - 334)
     INTEGRAL    Likelihood = -0.22   Transmembrane   41 -  57 (  41 -  57)

----- Final Results -----
              bacterial membrane --- Certainty=0.646(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 80/326 (24%), Positives = 157/326 (47%), Gaps = 34/326 (10%)

Query: 10  LLILLILLIAAIILFLIYGIPTDANEFL----------IIYILKTRYQKLIALILVGICI 59
            +L++L LL  A+I   + G+       +          +  I   R+ +++  +L G  I
Sbjct: 34  VLLILSLLFLAVIALSLGGLAVSYGAIVKGLFVAYDPQVALIYDLRFPRIVIALLAGAGI 93

Query: 60  GSSSLIFQTLTNNRLLTPSIIGL---DSLYILIQTGLMYLIGAQRVIKFSSFSSFL---L 113
              S ++FQ +  N +  P+IIG+    S  +L+ + L+      +++ +    SFL   +
Sbjct: 94  AVSGVLFQAVLKNPISDPAIIGICSGASFMVLVSSLLL-----PQLLLYGPIVSFLGGGV 148

Query: 114 SLLLMVGFAYLLFTILFRNKKQSLYFVLLAGLIFNTLFSSISSFIQAIMDPNDFMILQNQ 173
            S LL+ G A+         K  +   ++L G+  N LF  +S+ + +       M+  N
Sbjct: 149 SFLLIYGLAW--------KKGLNPIRLILTGIAINALFMGLSTALTSFFTSASPMV--NA 198

Query: 174 LFASFNAINTKI-LWISFIIIVVSFVINWPFIKELDVLLLGKENAISLGISYQKLTTRFF 232
            L A   +  T   + + F    +  ++     K  ++LLL  +    LGI    L
Sbjct: 199 LLAGHISQKTWADVGVLFPYTFIGLLLALLLSKTCNLLLLDDQVIRHLGIDATALRLGIS 258

Query: 233 LWLALMVAIATALVGPITFLGLLVAHITYHSFHTFRHQILVPIAIVICIFTLVLGQHLVQ 292
            L   L+ ++AT++VG ++FLGL+V H++     + +HQIL+P + ++  F  +L   L +
```

```
Sbjct: 259 LVAVLLASVATSIVGVVSFLGLIVPHMSRLLVGS-KHQILIPFSALLGAFVFLLADTLGR 317

Query: 293 NLLH-LTVQLSVLLNLIGGSYFIFTL 317
           +L + L +  +++++++GG YFI+ L
Sbjct: 318 SLAYPLEISPAIIMSIVGGPYFIYLL 343
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2491> which encodes amino acid sequence <SEQ ID 2492>. An alignment of the GAS and GBS sequences follows:

```
 Score = 51.9 bits (122), Expect = 5e-08
 Identities = 73/327 (22%), Positives = 137/327 (41%), Gaps = 38/327 (11%)

Query: 494 IISSLGTAISTVAQGIGTGLAIAFRGLGAAIAMVPPTTWLALGTAILMVGAAFALAGTQA 553
           +I  L T    +  G    L IA   +GA + +V      A+    L++ A
Sbjct: 573 VILGLVTTAVMMLLGAIAPLVIAIGAIGAPVGIVVAAIVGAIAVITLIIQAIMNWGA--- 629

Query: 554 DGISQILRTIGDXXXXXXXXXTDSLATLLTIIANAIGSMLPIVAGAISQIVG-------A 606
             I++ L++  D           ++ T  T    A  +    ++G  S +V        +
Sbjct: 630 --ITEWLQSTWDSCAAWXSELWTNIVTTAT---TAWSNFTAWLSGLWSSVVSTGQSLWSS 684

Query: 607 VAGGLSQLIIAVSTGVSLVIGAFTGLLGGI-SGVINSISAVIQSLTGVITAVFNGIATVI 665
               LS +  ++ TG   +  +FT  L  + SG++++ S +  +L+  I+ +FNGI +
Sbjct: 685 FTSSLSNIFSSLITGAQSLWSSFTSTLSNLWSGLVSTGSNLFNNLSSTISGIFNGILSTA 744

Query: 666 SSVGSTIKDVLTGLGTAFEGFGNGVKSALEGVGAVIESFGSAVR--------NVLDGVAN 717
           S++ ++IK  ++    A +G  N V +   GV A+   F   ++          + G AN
Sbjct: 745 SNIWNSIKSTIS---NAIDGAKNAVSN---GVNAIKNLFNFQIKWPHIPLPHFRVSGSAN 798

Query: 718 ILDSM--GTAALNAGRGVKEMAKGIKMLVDLSLGDLVATLAAVASGLGKMASSAGEMTTL 775
            LD +  G  ++    G+   AKG  ++   +L  +    A V    G  A      +TL
Sbjct: 799 PLDWLKGGLPSI----GIDWYAKG-GIMTKPTLFGMNGNRAMVGGEAGAEAILPLNKSTL 853

Query: 776 GSAMSKVANGMTRLATSATIAITGLTV 802
           G+    +AN M   + +  +  +G+T+
Sbjct: 854 GAIGQSIANTM-NTSNNINVNFSGVTI 879
 Score = 33.2 bits (74), Expect = 0.019
 Identities = 83/477 (17%), Positives = 175/477 (36%), Gaps = 103/477 (21%)

Query: 420 GSFLDKISTKFGLFGKKAKEGTD--------------QAANGSRKSGGIISQIFNGLGNI 465
           G  + +++T+FGL G+K K  ++               +A  ++++          LG +
Sbjct: 313 GDAVGELNTQFGLTGEKLKSASELLIKYAEINETDISSSAISAKQAIEAYGLTAEDLGMV 372

Query: 466 VKSAGTAISTAAKGIGTGIKTALSGAPPIISSLGTAISTVA--------QGIGTGLAIA- 516
           + +   A     + + T ++ A+ GAP  I  LG +    A         G+ +  A++
Sbjct: 373 LDNVTKAAQDTGQSVDTIVQKAIDGAPQ-IKGLGLSFEEGAALIGKFEKSGVDSSAALSS 431

Query: 517 ---------------FRGLGAAIAMVPPTT--WLALGTAILMVGAAFALAGTQA------ 553
                            GL   ++ +  +T    AL  A  + G+  A     A
Sbjct: 432 LSKAAVIYAKDGKTLTDGLNETVSAIQNSTSETEALSIASEIFGSKAAPRMVDAIQRGAF 491

Query: 554 --DGISQILRTIGDXXXXXXXXXTDSLATLLTI-------IANAIGSMLPIVAGAISQIV 604
             D +++  ++             D +  L          +A   G +L  V  A+  ++
Sbjct: 492 SFDDLAEAAKSSSGTVSTTFDETLDPIDKLTQYSNQAKEGMAELGGKLLETVIPALEPLM 551

Query: 605 GAVAGGLS----------QLII---AVSTGVSLVIGAFTGL---LGGISGVINSISAVIQ 648
           G +   ++          Q I+    V+T V +++GA   L   +G I   +  + A I
Sbjct: 552 GMLESSVNWFTSLNETDQQTIVILGLVTTAVMMLLGAIAPLVIAIGAIGAPVGIVVAAIV 611

Query: 649 SLTGVITAVFNGI-----------------ATVISSVGSTIKDVLTGLGTAFEGFGNGVK 691
               VIT +   I                 A   S + + I    T   + F  + +G+
Sbjct: 612 GAIAVITLIIQAIMNWGAITEWLQSTWDSCAAWXSELWTNIVTTATTAWSNFTAWLSGLW 671

Query: 692 SALEGVG-AVIESFGSAVRNV----LDGVANILDSMGTAALNAGRGVKEMAKGIKMLVDL 746
           S++   G ++  SF S++ N+    + G  ++  S  +   N   G+      +
Sbjct: 672 SSVVSTGQSLWSSFTSSLSNIFSSLITGAQSLWSSFTSTLSNLWSGLVSTGSNL------ 725

Query: 747 SLGDLVATLAAVASGLGKMASSAGEMTTLGSAMSKVANGMTRLATSATIAITGLTVF 803
```

```
                +L +T++ + +G+  +++++    ++ S +S   +G     ++   AI  L  F
Sbjct: 726 -FNNLSSTISGIFNGI--LSTASNIWNSIKSTISNAIDGAKNAVSNGVNAIKNLFNF 779
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 622

A DNA sequence (GBSx0662) was identified in *S.agalactiae* <SEQ ID 1935> which encodes the amino acid sequence <SEQ ID 1936>. Analysis of this protein sequence reveals the following:

```
Possible site: 13
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2277(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 623

A DNA sequence (GBSx0663) was identified in *S.agalactiae* <SEQ ID 1937> which encodes the amino acid sequence <SEQ ID 1938>. This protein is predicted to be membrane protein (ceuB). Analysis of this protein sequence reveals the following:

```
Possible site: 41
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-11.30   Transmembrane  241 - 257 ( 237 - 274)
   INTEGRAL    Likelihood = -6.42   Transmembrane  127 - 143 ( 118 - 149)
   INTEGRAL    Likelihood = -5.79   Transmembrane  152 - 168 ( 150 - 174)
   INTEGRAL    Likelihood = -5.47   Transmembrane  312 - 328 ( 309 - 330)
   INTEGRAL    Likelihood = -4.83   Transmembrane  289 - 305 ( 287 - 308)
   INTEGRAL    Likelihood = -4.67   Transmembrane   24 -  40 (  22 -  46)
   INTEGRAL    Likelihood = -4.35   Transmembrane   69 -  85 (  68 -  86)
   INTEGRAL    Likelihood = -4.19   Transmembrane  200 - 216 ( 198 - 216)
   INTEGRAL    Likelihood = -2.76   Transmembrane  107 - 123 ( 107 - 123)
   INTEGRAL    Likelihood = -0.85   Transmembrane  258 - 274 ( 258 - 274)

----- Final Results -----
               bacterial membrane --- Certainty=0.5522(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 8621> which encodes amino acid sequence <SEQ ID 8622> was also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 2
SRCFLG: 0
McG: Length  of  UR:    23
     Peak Value of UR:   2.64
     Net Charge of CR: 2
McG: Discrim Score:      8.59
GvH: Signal Score (-7.5): -4.6
     Possible site: 26
>>> Seems to have an uncleavable N-term signal seq
```

```
Amino Acid Composition: calculated from 1
ALOM program   count: 9 value: -11.30 threshold:  0.0
   INTEGRAL    Likelihood =-11.30   Transmembrane  226 - 242 ( 222 - 259)
   INTEGRAL    Likelihood = -6.42   Transmembrane  112 - 128 ( 103 - 134)
   INTEGRAL    Likelihood = -5.79   Transmembrane  137 - 153 ( 135 - 159)
   INTEGRAL    Likelihood = -4.67   Transmembrane    9 -  25 (   7 -  31)
   INTEGRAL    Likelihood = -4.35   Transmembrane   54 -  70 (  53 -  71)
   INTEGRAL    Likelihood = -4.19   Transmembrane  185 - 201 ( 183 - 201)
   INTEGRAL    Likelihood = -3.08   Transmembrane  268 - 284 ( 265 - 284)
   INTEGRAL    Likelihood = -2.76   Transmembrane   92 - 108 (  92 - 108)
   INTEGRAL    Likelihood = -0.85   Transmembrane  243 - 259 ( 243 - 259)
   PERIPHERAL  Likelihood =  5.73     203
 modified ALOM score:   2.76
icm1 HYPID: 7  CFP: 0.552

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.5522(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB12188 GB:Z99106 similar to ferrichrome ABC transporter
           (permease) [Bacillus subtilis]
 Identities = 149/304 (49%), Positives = 234/304 (76%)

Query: 29  LVILSLTSLFVGVKSIPLEQITHLDQSQVDIFLTSRLPRTISILISGASLSVCGLLMQQL 88
            L+IL++TS+F+GV+ +    +  L + +      SRLPR ISI+I+G S+S+CGL+MQQ+
Sbjct: 10  LIILAVTSVFIGVEDLSPLDLFDLSKQEASTLFASRLPRLISIVIAGLSMSICGLIMQQI 69

Query: 89  TQNKFVSPTTSGTMDWAKLGVVVTLIFFKNTSIFIQLCIASGFAILGSLLFVTILKMITF 148
            ++NKFVSPTT+GTMDWA+LG++++L+ F + S  I++ +A  FA+ G+ LF+ IL+ I F
Sbjct: 70  SRNKFVSPTTAGTMDWARLGILISLLLFTSASPLIKMLVAFVFALAGNFLFMKILERIKF 129

Query: 149 KDNIFIPLIGLMLGQIVAAATVFLGTHFQVLQSVNSWLQGNFSIMTSHRYEILYLALPCL 208
             D IFIPL+GLMLG IV++   F+   + ++Q+V+SWLQG+FS++   RYE+LYL++P +
Sbjct: 130 NDTIFIPLVGLMLGNIVSSIATFIAYKYDLIQNVSSWLQGDFSLVVKGRYELLYLSIPLV 189

Query: 209 FLVYFFAHQFTIVGLGESFAKNLGVAYEKMIYFGLVLVSIMTSLVIIIVGALPFLGLIVP 268
             + Y +A +FT+ G+GESF+ NLG+ Y++++  GL++VS++TSLVI+ VG LPFLGLI+P
Sbjct: 190 IIAYVYADKFTLAGMGESFSVNLGLKYKRVVNIGLIIVSLITSLVILTVGMLPFLGLIIP 249

Query: 269 NLISITKGDHMSSTILETSLLGACIVMICDLFGRLVIFPYEVSIGVTLGVLGSAFFLISI 328
            N++SI +GD++ S++  T LLGA  V+ CD+ GR++IFPYE+SIG+ +G++GS  FL  +
Sbjct: 250 NIVSIYRGDNLKSSLPHTVLLGAVFVLFCDILGRIIIFPYEISIGLMVGIIGSGIFLFML 309

Query: 329 IRNE 332
            +R +
Sbjct: 310 LRRK 313
```

There is also homology to SEQ ID 1940.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 624

A DNA sequence (GBSx0664) was identified in *S.agalactiae* <SEQ ID 1941> which encodes the amino acid sequence <SEQ ID 1942>. Analysis of this protein sequence reveals the following:

```
Possible site: 35
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.90   Transmembrane  140 - 156 ( 140 - 156)
```

```
----- Final Results -----
               bacterial membrane --- Certainty=0.1362(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAB06720 GB:AP001517 maltose transacetylase (maltose
           O-acetyltransferase) [Bacillus halodurans]
 Identities = 93/182 (51%), Positives = 125/182 (68%), Gaps = 2/182 (1%)

Query: 2   TEKEKMLAGQYYRPSAPELRKDREVALKNMQAFNN--EDNSSKRNVILQKWFGATGKSIH 59
            TEKEKMLAG+ Y+   PEL KDRE A +  + FN   E    +R  ++++ FG+ G+S++
Sbjct: 3   TEKEKMLAGERYKAWDPELVKDRERARRLTRLFNQTTETEEKQRTELIKELFGSMGESVN 62

Query: 60  MEQRFVCDYGCNIYVGENFYANFNQTFLDVCEIRIGDNCMFGPNCQLLTPLHPLDPIERN 119
            +E  F CDYG NI+VG NF+ANF+   LDVCE+RIG NCM  P   + T  HP+ P+ER
Sbjct: 63  IEPTFRCDYGYNIHVGNNFFANFDCVILDVCEVRIGANCMLAPGVHIYTATHPIHPLERV 122

Query: 120 SGLEYGAPIQIGNNVWLGGGVTILPGVVLGDNVVVGAGSVVTKSFENNVVIAGNPAKIIKKL 182
             G EYG P+ I NNVW+GG   + PGV +G+N V+ +GSVVTK    NVV+AGNPAK+I+ +
Sbjct: 123 EGPEYGKPVTIRNNVWIGGRAIVNPGVTIGNNAVIASGSVVTKDVPENVVVAGNPAKVIQTI 184
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1943> which encodes the amino acid sequence <SEQ ID 1944>. Analysis of this protein sequence reveals the following:

```
Possible site: 61
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4052(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 68/188 (36%), Positives = 101/188 (53%), Gaps = 13/188 (6%)

Query: 2   TEKEKMLAGQYYRPSAPELRKDREVALKNMQAFN--------NEDNSSKRNVILQKWFGA 53
            TE +KM  G++Y     +   D E+  K M A +              +R+ +L + FG
Sbjct: 3   TEFDKMTRGEWY-----DANFDSELIQKRMMAQDLCFDLNQLKPSREEERSAVLNQLFGQ 57

Query: 54  TGKSIHMEQRFVCDYGCNIYVGENFYANFNQTFLDVCEIRIGDNCMFGPNCQLLTPLHPL 113
            + + + +   F+CDYG NI  G+N + N N  F+D  +I +GDN   GP+    T  HPL
Sbjct: 58  SFEGLVLLSPFICDYGKNITFGKNCFINSNCYFMDGAKIALGDNVFVGPSTGFYTANHPL 117

Query: 114 DPIERNSGLEYGAPIQIGNNVWLGGGVTILPGVVLGDNVVVGAGSVVTKSFENNVVIAGN 173
            D   RN GLE   PI IG+NVW G  V ++PGV +G   V+ +GSVVT     N + AG
Sbjct: 118 DYKRRNEGLEKALPITIGDNVWFGANVNVMPGVTIGSGCVIASGSVVTHDIPVNSLAAGV 177

Query: 174 PAKIIKKL 181
            P ++++K+
Sbjct: 178 PCQVVRKI 185
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 625

A DNA sequence (GBSx0665) was identified in *S.agalactiae* <SEQ ID 1945> which encodes the amino acid sequence <SEQ ID 1946>. This protein is predicted to be ribonuclease H (rnhB-2). Analysis of this protein sequence reveals the following:

```
Possible site: 32
```

```
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.16   Transmembrane   79 -  95 (  79 -  95)

----- Final Results -----
               bacterial membrane --- Certainty=0.1065(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9823> which encodes amino acid sequence <SEQ ID 9824> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB13479 GB:Z99112 ribonuclease H [Bacillus subtilis]
 Identities = 128/249 (51%), Positives = 168/249 (67%)

Query: 4   TIKEIKAILETIVDLKDKRWQEYQTDSRAGVQKAILQRKKNIQSDLDEEARLEQMLVYEK 63
           T+K+IK  L+ + D +D    + + D R  VQ  + Q  K    +   + +   M  YE+
Sbjct: 5   TVKDIKDRLQEVKDAQDPFIAQCENDPRKSVQTLVEQWLKKQAKEKALKEQWVNMTSYER 64

Query: 64  KLYIEHINLIAGIDEVGRGPLAGPVVAAAVILPPNCKIKHLNDSKKIPKKKHQEIYQNIL 123
               +   LIAG+DEVGRGPLAGPVVA+AVILP  C+I  L DSKK+ +KK +E Y+ I+
Sbjct: 65  LARNKGFRLIAGVDEVGRGPLAGPVVASAVILPEECEILGLTDSKKLSEKKREEYYELIM 124

Query: 124 DQALAVGIGIQDSQCIDDINIYEATKHAMIDAVSHLSVAPEHLLIDAMVLDLSIPQTKII 183
            +ALAVGIGI ++  ID+INIYEA+K AM+ A+  LS  P++LL+DAM L L   Q  II
Sbjct: 125 KEALAVGIGIVEATVIDEINIYEASKMAMVKAIQDLSDTPDYLLVDAMTLPLDTAQASII 184

Query: 184 KGDANSLSIAAASIVAKVTRDKIMSDYDSTYPGYAFSKNAGYGTKEHLEGLQKYGITPIH 243
           KGDA S+SIAA + +AKVTRD++MS Y  TYP Y F KN GYGTKEHLE L  YG T +H
Sbjct: 185 KGDAKSVSIAAGACIAKVTRDRMMSAYAETYPMYGFEKNKGYGTKEHLEALAAYGPTELH 244

Query: 244 RKSFEPIKS 252
           RK+F P++S
Sbjct: 245 RKTFAPVQS 253
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1947> which encodes the amino acid sequence <SEQ ID 1948>. Analysis of this protein sequence reveals the following:

```
     Possible site: 50

>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.53   Transmembrane   79 -  95 (  79 -  95)

----- Final Results -----
               bacterial membrane --- Certainty=0.1213(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:CAB13479 GB:Z99112 ribonuclease H [Bacillus subtilis]
 Identities = 130/252 (51%), Positives = 176/252 (69%), Gaps = 3/252 (1%)

Query: 4   SIKAIKESLEAVTSLLDPLFQELATDTRSGVQKALKSRQKVIQAELAEEERLEAMLSYEK 63
           ++K IK+ L+ V    DP   +   D R  VQ  ++   K    E A +E+   M SYE+
Sbjct: 5   TVKDIKDRLQEVKDAQDPFIAQCENDPRKSVQTLVEQWLKKQAKEKALKEQWVNMTSYER 64

Query: 64  ALYKKGYKAIAGIDEVGRGPLAGPVVAACVILPKYCKIKGLNDSKKIPKAKHETIYQAVK 123
               KG++ IAG+DEVGRGPLAGPVVA+ VILP+ C+I GL DSKK+ + K E  Y+ +
Sbjct: 65  LARNKGFRLIAGVDEVGRGPLAGPVVASAVILPEECEILGLTDSKKLSEKKREEYYELIM 124

Query: 124 EKALAIGIGIIDNQLIDEVNIYEATKLAMLEAIKQLEGQLTQPDYLLIDAMTLDIAISQQ 183
           ++ALA+GIGI++  +IDE+NIYEA+K+AM++AI+ L      PDYLL+DAMTL +  +Q
Sbjct: 125 KEALAVGIGIVEATVIDEINIYEASKMAMVKAIQDLS---DTPDYLLVDAMTLPLDTAQA 181
```

```
Query: 184 SILKGDANSLSIAAASIVAKVTRDQMMANYDRIFPGYDFAKNAGYGTKEHLQGLKAYGIT 243
           SI+KGDA S+SIAA + +AKVTRD+MM+ Y   +P Y F KN GYGTKEHL+ L AYG T
Sbjct: 182 SIIKGDAKSVSIAAGACIAKVTRDRMMSAYAETYPMYGFEKNKGYGTKEHLEALAAYGPT 241

Query: 244 PIHRKSFEPVKS 255
            +HRK+F PV+S
Sbjct: 242 ELHRKTFAPVQS 253
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 168/256 (65%), Positives = 203/256 (78%), Gaps = 3/256 (1%)

Query: 1   MMATIKEIKAILETIVDLKDKRWQEYQTDSRAGVQKAILQRKKNIQSDLDEEARLEQMLV 60
           M  +IK IK  LE +  L D  +QE  TD+R+GVQKA+  R+K IQ++L EE RLE ML
Sbjct: 1   MPTSIKAIKESLEAVTSLLDPLFQELATDTRSGVQKALKSRQKVIQAELAEEERLEAMLS 60

Query: 61  YEKKLYIEHINLIAGIDEVGRGPLAGPVVAAAVILPPNCKIKHLNDSKKIPKKKHQEIYQ 120
           YEK LY +    IAGIDEVGRGPLAGPVVAA VILP  CKIK LNDSKKIPK KH+ IYQ
Sbjct: 61  YEKALYKKGYKAIAGIDEVGRGPLAGPVVAACVILPKYCKIKGLNDSKKIPKAKHETIYQ 120

Query: 121 NILDQALAVGIGIQDSQCIDDINIYEATKHAMIDAVSHLS---VAPEHLLIDAMVLDLSI 177
            + ++ALA+GIGI D+Q ID++NIYEATK AM++A+  L      P++LLIDAM LD++I
Sbjct: 121 AVKEKALAIGIGIIDNQLIDEVNIYEATKLAMLEAIKQLEGQLTQPDYLLIDAMTLDIAI 180

Query: 178 PQTKIIKGDANSLSIAAASIVAKVTRDKIMSDYDSTYPGYAFSKNAGYGTKEHLEGLQKY 237
            Q  I+KGDANSLSIAAASIVAKVTRD++M++YD  +PGY F+KNAGYGTKEHL+GL+ Y
Sbjct: 181 SQQSILKGDANSLSIAAASIVAKVTRDQMMANYDRIFPGYDFAKNAGYGTKEHLQGLKAY 240

Query: 238 GITPIHRKSFEPIKSM 253
           GITPIHRKSFEP+KSM
Sbjct: 241 GITPIHRKSFEPVKSM 256
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 626

A DNA sequence (GBSx0666) was identified in *S.agalactiae* <SEQ ID 1949> which encodes the amino acid sequence <SEQ ID 1950>. Analysis of this protein sequence reveals the following:

```
Possible site: 16
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1865(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 627

A DNA sequence (GBSx0667) was identified in *S.agalactiae* <SEQ ID 1951> which encodes the amino acid sequence <SEQ ID 1952>. Analysis of this protein sequence reveals the following:

```
Possible site: 14
>>> Seems to have no N-terminal signal sequence
```

```
----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3034(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAB06195 GB:AP001515 unknown conserved protein [Bacillus halodurans]
 Identities = 140/281 (49%), Positives = 195/281 (68%), Gaps = 5/281 (1%)

Query: 3   TIQWFPGHMSKARRQVQENIKHVDFVTILVDARLPLSSQNPMLTKIVGDKPKLMILNKAD 62
            TIQWFPGHM+KARR+V E +K +D V  L+DAR+PLSS+NPM+ +IV  KP+L++LNK D
Sbjct: 2   TIQWFPGHMAKARREVTEKLKLIDVVIELLDARVPLSSRNPMMDEIVAHKPRLVLLNKDD 61

Query: 63  LADPIRTKEWRDFYESQGLKTLAINSKEQSTVKKVTDIAKILMSDKIANLRGRGIQKETL 122
            LADP +TKEW  F+E  G   L IN++    V +++   + L    I   R +G++   +
Sbjct: 62  LADPSKTKEWTRFFEEGGATVLPINAQTGQGVSRISPACQTLAQALIEKQRAKGMKPRAI 121

Query: 123 RTMIIGIPNAGKSTLMNRLAGKKIAVVGNKPGVTKGQQWLKSNKELEILDTPGILWPKFE 182
            R MI+GIPN GKSTL+NRLA K+IA VG++PG+TK QQW+K  KELE+LDTPGILWPKF+
Sbjct: 122 RAMILGIPNVGKSTLINRLASKRIAKVGDRPGITKQQQWIKVGKELELLDTPGILWPKFD 181

Query: 183 DELVGLKLALTGAIKDQLLPMDEVTIFGLNYFKTYYPDRLKERFKSINLEDEAPEIIMAL 242
            D+  G +LA TGAIKD+LL   +V +F L Y +  YPDRL +R+K   L ++   +  A+
Sbjct: 182 DQATGFRLAATGAIKDELLDFQDVALFVLRYMREMYPDRLMDRYKLNELPEDGVTLFDAI 241

Query: 243 TQKLGY-----RDDYDRFYNLFVKEVRDGKLGRYTLDIVGE 278
             +K G+       DYD+   + ++E+R G LGR TL++ G+
Sbjct: 242 GKKRGHLLSGGYIDYDKTAEMILRELRAGTLGRITLEVPGK 282
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1953> which encodes the amino acid sequence <SEQ ID 1954>. Analysis of this protein sequence reveals the following:

```
Possible site: 14
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2688(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 247/282 (87%), Positives = 265/282 (93%)

Query: 1   MATIQWFPGHMSKARRQVQENIKHVDFVTILVDARLPLSSQNPMLTKIVGDKPKLMILNK 60
            MA IQWFPGHMSKARRQVQEN+KHVDFVTILVDARLPLSSQNPMLTKIVGDKPKLMILNK
Sbjct: 1   MAMIQWFPGHMSKARRQVQENVKHVDFVTILVDARLPLSSQNPMLTKIVGDKPKLMILNK 60

Query: 61  ADLADPIRTKEWRDFYESQGLKTLAINSKEQSTVKKVTDIAKILMSDKIANLRGRGIQKE 120
            ADLAD  RTKEW+ +YESQG+KTLAINSKEQSTVKKVT+ AK LM+DKI  LR RGIQKE
Sbjct: 61  ADLADATRTKEWKAYYESQGIKTLAINSKEQSTVKKVTEAAKELMADKIQRLRERGIQKE 120

Query: 121 TLRTMIIGIPNAGKSTLMNRLAGKKIAVVGNKPGVTKGQQWLKSNKELEILDTPGILWPK 180
            TLRTMIIGIPNAGKSTLMNRLAGKKIAVVGNKPGVTKGQQWLKSNKELEILDTPGILWPK
Sbjct: 121 TLRTMIIGIPNAGKSTLMNRLAGKKIAVVGNKPGVTKGQQWLKSNKELEILDTPGILWPK 180

Query: 181 FEDELVGLKLALTGAIKDQLLPMDEVTIFGLNYFKTYYPDRLKERFKSINLEDEAPEIIM 240
            FEDELVGLKLALTGAIKDQLLPMDEVTIFGLNYF+ YYP+RL +RFK+I LE+EAPEIIM
Sbjct: 181 FEDELVGLKLALTGAIKDQLLPMDEVTIFGLNYFREYYPNRLTKRFKNIPLEEEAPEIIM 240

Query: 241 ALTQKLGYRDDYDRFYNLFVKEVRDGKLGRYTLDIVGEHDGN 282
             LT++LG++DDYDRFY LFVKEVRDGKLG+YTLD VG+ D +
Sbjct: 241 TLTRQLGFKDDYDRFYTLFVKEVRDGKLGQYTLDQVGDMDAD 282
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 628

A DNA sequence (GBSx0668) was identified in *S.agalactiae* <SEQ ID 1955> which encodes the amino acid sequence <SEQ ID 1956>. Analysis of this protein sequence reveals the following:

```
Possible site: 24
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
               bacterial outside --- Certainty=0.3000(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9825> which encodes amino acid sequence <SEQ ID 9826> was also identified.

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB12129 GB:Z99105 similar to hypothetical proteins [Bacillus subtilis]
 Identities = 69/173 (39%), Positives = 102/173 (58%), Gaps = 13/173 (7%)

Query: 29  DKAKEKASV-----IKQASQTSQTSKKEVLQKKT----YPNLNKYSNLEIHVSSTRQTMT 79
            D A+E AS+     ++  +   +T+K  +   K     YP++ K  ++ I V+   Q
Sbjct: 22  DHAEEHASINTKKTVENITDVRKTAKTSIDWTKPSGGEYPDI-KQKHVWIDVNVKEQKAY 80

Query: 80  ITSNDKVIFKTIVSTG---AKESPTPKGTFVIEPERGDFFYNASSKEGAYYWVSFKEHGI 136
            I      I+  ++S+G    K+  TPKGTF +EPERG++F++   +EGA YWVS+K HG
Sbjct: 81  IKEGSNTIYTMMISSGLDQTKDDATPKGTFYVEPERGEWFFSEGYQEGAEYWVSWKNHGE 140

Query: 137 YLFHSVPTDQQGNEIPEEAKQLGKAASHGCVRMSRADAKWFYENIPQGTTVTI 189
            +LFHSVP  +    I  EA++LG   SHGC+R++  DAKW YENIP+ T V I
Sbjct: 141 FLFHSVPMTKDQKVIKTEAEKLGTKVSHGCIRLTIPDAKWVYENIPEHTKVVI 193
```

No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 1956 (GBS644) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 130 (lane 2 & 3; MW 49.6kDa) and in Figure 186 (lane 3; MW 50kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 130 (lane 5-7; MW 24.6kDa) and in Figure 177 (lane 3; MW 25kDa).

GBS644-GST was purified as shown in Figure 236, lane 9.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 629

A DNA sequence (GBSx0669) was identified in *S.agalactiae* <SEQ ID 1957> which encodes the amino acid sequence <SEQ ID 1958>. This protein is predicted to be carbon starvation protein A. Analysis of this protein sequence reveals the following:

```
Possible site: 19
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-11.25   Transmembrane  129 - 145 ( 122 - 157)
   INTEGRAL    Likelihood = -9.92   Transmembrane  316 - 332 ( 305 - 342)
   INTEGRAL    Likelihood = -6.42   Transmembrane  164 - 180 ( 157 - 181)
   INTEGRAL    Likelihood = -5.73   Transmembrane  443 - 459 ( 441 - 466)
```

```
   INTEGRAL    Likelihood = -5.57   Transmembrane  416 - 432 ( 414 - 435)
   INTEGRAL    Likelihood = -4.88   Transmembrane  190 - 206 ( 183 - 209)
   INTEGRAL    Likelihood = -4.83   Transmembrane   78 -  94 (  70 -  95)
   INTEGRAL    Likelihood = -3.13   Transmembrane  362 - 378 ( 359 - 379)
   INTEGRAL    Likelihood = -2.34   Transmembrane  228 - 244 ( 227 - 245)
   INTEGRAL    Likelihood = -2.02   Transmembrane    2 -  18 (   1 -  18)
   INTEGRAL    Likelihood = -1.28   Transmembrane  393 - 409 ( 393 - 410)

----- Final Results -----
              bacterial membrane --- Certainty=0.5501(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAF93852 GB:AE004154 carbon starvation protein A, putative
          [Vibrio cholerae]
 Identities = 220/470 (46%), Positives = 311/470 (65%), Gaps = 16/470 (3%)

Query: 1   MVTFLGGVALLIVGYFTYGRYIEKNFQIDENRQTPAEALRDGYDFVPMPKWKNGMIELLN 60
            M+ FL  VA L+ GYF YG ++EK F I+E RQTPA    DG D+VPM   K  +++LLN
Sbjct: 1   MLWFLTCVAALVGGYFIYGAFVEKVFGINEKRQTPAHTKTDGVDYVPMSTPKVYLVQLLN 60

Query: 61  IAGTGPIFGPILGALYGPVAYIWIVLGCIFAGAVHDYMIGMISLRNNGAYLPELASRYLG 120
            IAG GPIFGPI+GALYGP A +WIV+GCIFAGAVHDY  GM+S+RN GA +P +  RYLG
Sbjct: 61  IAGVGPIFGPIMGALYGPAAMLWIVVGCIFAGAVHDYFSGMLSIRNGGASVPSITGRYLG 120

Query: 121 KSMKHVINIFSMLLLILVATVFVVTPANLILSILPAG---TLSLPWIIGLIFVYYLISTV 177
               KH +NIF+++LL+LV  VFV  PA +I +++      T+S+  ++ +IF YY+++T+
Sbjct: 121 NGAKHFMNIFAIVLLLLVGVVFVSAPAGMITNLINQQTDFTVSMTTMVVIIFAYYILATI 180

Query: 178 LPIDKALGKVYPVF-------CVILMVSTAAVGFRLLTGGFDMPNLTFETFKNMHPAGLG 230
            +P+DK +G+ YP+F        V LM + A      + GGF++ ++     KN++P  +
Sbjct: 181 VPVDKIIGRFYPLFGALLIFMSVGLMTAIAFSSEHQVLGGFEISDMV----KNLNPNDMP 236

Query: 231 IFPALFFTISCGAISGFHATQAPMVSRTTVNEREGRFTFYGMMIAEGVIAMIWAGASMSL 290
            ++PALF TI+CGAISGFHATQ+P+++R   NE+ GRF FYG MI EG+IA+IW   ++S
Sbjct: 237 LWPALFITIACGAISGFHATQSPLMARCMENEKNGRFVFYGAMIGEGIIALIWCTVALSF 296

Query: 291 FKG-QNLYEMIAAGTPSAVVNQVMLMLLGSVIGTIAIIGVIVLPVSSGLSAFRSLRTIVA 349
            F   + L E +  G P  VV      LLG   G IA +GV++LP++SG +AFRS R I+A
Sbjct: 297 FGSLEALSEAVKNGGPGNVVYGASFGLLGVFGGVIAFLGVVILPITSGDTAFRSSRLILA 356

Query: 350 DYIHVKQDTLPKIFAVTIPLYVISFVLTHVDFNLLWRYFNWANQVTAVIGLLVATRYLIL 409
            +Y +++Q TL     + +PL+VI  VLT VDF ++WRYF +ANQ TAV+ L  AT YL+
Sbjct: 357 EYFNMEQKTLRNRLLMAVPLFVIGAVLTQVDFGIIWRYFGFANQATAVMMLWTATAYLMR 416

Query: 410 KRRNYWVTFVPAMFMLYAVVVYIL-SQPIGFNMGLGILTYSLALVLTGIL 458
              + +W+  VPA+FM    + +IL S  +GF + + I T +  L   G L
Sbjct: 417 HNKLHWICTVPALFMTTVCISFILNSSTLGFGLPMQISTIAGVLASLGAL 466
```

No corresponding DNA sequence was identified in *S.pyogenes.*

A related GBS gene <SEQ ID 8623> and protein <SEQ ID 8624> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 0
McG: Discrim Score:      6.07
GvH: Signal Score (-7.5): -3.54
     Possible site: 19
>>> Seems to have an uncleavable N-term signal seq
ALOM program   count: 11 value: -11.25 threshold:  0.0
   INTEGRAL    Likelihood =-11.25   Transmembrane  129 - 145 ( 122 - 157)
   INTEGRAL    Likelihood = -9.92   Transmembrane  316 - 332 ( 305 - 342)
   INTEGRAL    Likelihood = -6.42   Transmembrane  164 - 180 ( 157 - 181)
   INTEGRAL    Likelihood = -5.57   Transmembrane  416 - 432 ( 414 - 435)
   INTEGRAL    Likelihood = -4.88   Transmembrane  190 - 206 ( 183 - 209)
```

```
   INTEGRAL    Likelihood = -4.83   Transmembrane   78 -  94 (  70 -  95)
   INTEGRAL    Likelihood = -4.67   Transmembrane  445 - 461 ( 441 - 463)
   INTEGRAL    Likelihood = -3.13   Transmembrane  362 - 378 ( 359 - 379)
   INTEGRAL    Likelihood = -2.34   Transmembrane  228 - 244 ( 227 - 245)
   INTEGRAL    Likelihood = -2.02   Transmembrane    2 -  18 (   1 -  18)
   INTEGRAL    Likelihood = -1.28   Transmembrane  393 - 409 ( 393 - 410)
   PERIPHERAL  Likelihood =  0.21     272
 modified ALOM score:   2.75

*** Reasoning Step: 3

----- Final Results -----
               bacterial membrane --- Certainty=0.5501(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
ORF01729(301 - 1668 of 2082)
GP|9655126|gb|AAF93852.1||AE004154(1 - 464 of 494) carbon starvation protein A, putative
{Vibrio cholerae}
%Match = 29.9
%Identity = 47.6  %Similarity = 68.6
Matches = 218  Mismatches = 138  Conservative Sub.s = 96

174       204       234       264       294       324       354       384
TNEKLFIIKLRLFISKKQPFILKIGNFNFSMLY*SHENA**N*AKKFLGGSDMVTFLGGVALLIVGYFTYGRYIEKNFQI
                                                    |: ||  || |: ||| || ::|| | |
                                                    MLWFLTCVAALVGGYFIYGAFVEKVFGI
                                                            10        20

414       444       474       504       534       564       594       624
DENRQTPAEALRDGYDFVPMPKWKNGMIELLNIAGTGPIFGPILGALYGPVAYIWIVLGCIFAGAVHDYMIGMISLRNNG
:| |||||    || |:|||   |  :::|||||| |||||||:|||||| | :|||:|||||||||||  ||:|:|| |
NEKRQTPAHTKTDGVDYVPMSTPKVYLVQLLNIAGVGPIFGPIMGALYGPAAMLWIVVGCIFAGAVHDYFSGMLSIRNGG
          40        50        60        70        80        90       100

654       684       714       744       774       795       825       855
AYLPELASRYLGKSMKHVINIFSMLLLILVATVFVVTPANLILSILPAGT---LSLPWIIGLIFVYYLISTVLPIDKALG
| :| :  ||||   || :|||:::||:||  |||  || :| :::   |   :|:  :: :|| ||:::|::|:|| :|
ASVPSITGRYLGNGAKHFMNIFAIVLLLLVGVVFVSAPAGMITNLINQQTDFTVSMTTMVVIIFAYYILATIVPVDKIIG
         120       130       140       150       160       170       180

894       924       954       984       1014      1044      1074
KVYP------VFCVI-LMVSTAAVGFRLLTGGFDMPNLTFETFKNMHPAGLGIFPALFFTISCGAISGFHATQAPMVSRT
: ||      :|  : || : |      : |||:: ::     ||::|  : ::|||| ||:|||||||||||:|:::|
RFYPLFGALLIFMSVGLMTAIAFSSEHQVLGGFEISDM----VKNLNPNDMPLWPALFITIACGAISGFHATQSPLMARC
         200       210       220           230       240       250       260

1104      1134      1164      1191      1221      1251      1281      1311
TVNEREGRFTFYGMMIAEGVIAMIWAGASMSLFKG-QNLYEMIAAGTPSAVVNQVMLMLLGSVIGTIAIIGVIVLPVSSG
  ||: ||| ||| || ||:||:||   ::|:|   : | | :  | |  ||    : |||   | || :||::||::||
MENEKNGRFVFYGAMIGEGIIALIWCTVALSFFGSLEALSEAVKNGGPGNVVYGASFGLLGVFGGVIAFLGVVILPITSG
             280       290       300       310       320       330       340

1341      1371      1401      1431      1461      1491      1521      1551
LSAFRSLRTIVADYIHVKQDTLPKIFAVTIPLYVISFVLTHVDFNLLWRYFNWANQVTAVIGLLVATRYLILKRRNYWVT
 :|||| | |:|:| :::| ||   : : :||:||  |||:||| ::|||| :||| |||: |  || ||:   : :|:
DTAFRSSRLILAEYFNMEQKTLRNRLLMAVPLFVIGAVLTQVDFGIIWRYFGFANQATAVMMLWTATAYLMRHNKLHWIC
             360       370       380       390       400       410       420

1581      1608      1638      1668      1698      1728      1758      1788
FVPAMFMLYAVVVYIL-SQPIGFNMGLGILTYSLALVLTGIXVGLFWKSGQKQLKTVHPEAFLFNDHRPINYYSSLDS*Y
 |||:||    : :|| |  :|| : : | | :  |   |
TVPALFMTTVCISFILNSSTLGFGLPMQISTIAGVLASLGALAYVAKVSKGKGETDLADEEKPQGVTKTA
             440       450       460       470       480       490
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 630

A DNA sequence (GBSx0670) was identified in *S.agalactiae* <SEQ ID 1959> which encodes the amino acid sequence <SEQ ID 1960>. This protein is predicted to be lytR (lytT). Analysis of this protein sequence reveals the following:

```
Possible site: 30
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.80    Transmembrane   27 -  43 (  27 -  43)

----- Final Results -----
               bacterial membrane --- Certainty=0.1319(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAB48183 GB:L42945 lytR [Staphylococcus aureus]
 Identities = 93/245 (37%), Positives = 150/245 (60%), Gaps = 3/245 (1%)

Query: 1   MKVLVVDDEPVARNELIYLLNKYDSNLVIAEAHDMATALAILLRETFDVALLDIHLRDDS 60
           MK L++DDEP+ARNEL YLLN+      I EA ++   L  LL   +D+  LD++L D++
Sbjct: 1   MKALIIDDEPLARNELTYLLNEIGGFEEINEAENVKETLEALLINQYDIIFLDVNLMDEN 60

Query: 61  GLQLAEYINKMPKPPLLIFATAYDQYAIQAFEHDARDYLLKPYDFDRLKQAMDRVKGALS 120
           G++L   I KM +PP +IFATA+DQYA+QAFE +A DY+LKP+   R++QA+++V+    +
Sbjct: 61  GIELGAKIQKMKEPPAIIFATAHDQYAVQAFELNATDYILKPFGQKRIEQAVNKVRATKA 120

Query: 121 TSTIIESVTSGPL---FKQQYPLTVEDRIYLVSADDILLIEAMQGKLIIQTPDKNYEIDG 177
                 S+ +      F Q  P+ ++D+I+++   +I+ I    G   I T +  YE
Sbjct: 121 KDDNNASAIANDMSANFDQSLPVEIDDKIHMLKQQNIIGIGTHNGITTIHTTNHKYETTE 180

Query: 178 SLQQWQDKLPSSQFVRVHRSYIVNINAIKTIEPWFNQTLQLHLCNKITVPVSRANVKPLK 237
            L +++ +L  + F+R+HRSYI+N   IK ++ WFN T  + L N + + V R+ +K  K
Sbjct: 181 PLNRYEKRLNPTYFIRIHRSYIINTKHIKEVQQWFNYTYMVILTNGVKMQVGRSFMKDFK 240

Query: 238 QMLGI 242
             +G+
Sbjct: 241 ASIGL 245
```

There is also homology to SEQ ID 460.

SEQ ID 1960 (GBS399) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 75 (lane 7; MW 30.4kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 84 (lane 2; MW 55kDa). Purified GBS399-GST is shown in Figure 217, lane 9; purified GBS399d-GST is shown in Figure 236, lane 3.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 631

A DNA sequence (GBSx0671) was identified in *S.agalactiae* <SEQ ID 1961> which encodes the amino acid sequence <SEQ ID 1962>. Analysis of this protein sequence reveals the following:

```
Possible site: 51
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -7.59    Transmembrane   95 - 111 (  86 - 116)
```

```
     INTEGRAL    Likelihood = -5.95   Transmembrane  155 - 171 ( 152 - 176)
     INTEGRAL    Likelihood = -2.28   Transmembrane  189 - 205 ( 187 - 206)
     INTEGRAL    Likelihood = -1.49   Transmembrane  122 - 138 ( 121 - 138)

  ----- Final Results -----
              bacterial membrane --- Certainty=0.4036(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAB48182 GB:L42945 lytS [Staphylococcus aureus]
 Identities = 264/570 (46%), Positives = 389/570 (67%), Gaps = 2/570 (0%)

Query: 1   MTLFLIMMERAGLIILLAYAFVHIPFIKQTLKQPELKKHQYILLILFSLFAIISNFTGVE 60
           ++L ++++ER GLII+LAY  ++IP+ K  + +    K ++ L I+FSLFA++SN TG+
Sbjct: 2   LSLTMLLLERVGLIIILAYVLMNIPYFKNLMNRRRTWKARWQLCIIFSLFALMSNLTGIV 61

Query: 61  IQSDLSIIPQTLNHIADQSSVANTRVLTIGVSGLIGGPIVGIIVGLLSVFVRYLQGGLAP 120
           I    S+       + D  S+ANTRVLTIGV+GL+GGP VG+ VG++S   R   GG
Sbjct: 62  IDHQHSLSGSVYFRLDDDVSLANTRVLTIGVAGLVGGPFVGLFVGVISGIFRVYMGGADA 121

Query: 121 HIYVISSLLIGLCSGLSGNYLRKNYNKIRVLDAMVVGFGMEILQMICILIFSVDFNQALR 180
            +Y+ISS+ IG+ +G  G   ++      +  + ++G  ME++QM+ IL FS D   A+
Sbjct: 122 QVYLISSIFIGIIAGYFGLQAQRRKRYPSIAKSAMIGIVMEMIQMLSILTFSHDKAYAVD 181

Query: 181 LVSFISMPMILSNTLGLGIFISIISSTQKLEEHAKAFQTHQVLELANLTLPYLRKGLTTE 240
           L+S I++PMI+ N++G  IF+SII  T K E+  K  QTH VL+L N T PY ++GL  E
Sbjct: 182 LISLIALPMIIVNSVGPAIFMSIIIPTLKQEDQMKPVQTHDVLQLMNQTFPYFKEGLNRE 241

Query: 241 SCQPVAEIIHKHMDVSAVSLTSQSAILAYVGDGADHHLPNTQILTKLAKRAIDTGKVSVA 300
           S Q +A II   M VS+V++TS++ IL++VG G+DHH+P  +ILT L+K  + +GK+
Sbjct: 242 SAQQIAMIIKNLMKVSSVAITSKNEILSHVGGGSDHHIPTNEILTSLSKDVLKSGKLKEV 301

Query: 301 TDKSEIECDHKNCPLSSAIVIPLHIHDVIVGTLKLYFSDAQHMTYVDRQLAEGLGNIFST 360
             K EI C H NCPL +AIVIPL +H  IVGTLK+YF++   +T+V+RQLAEGL NIFS+
Sbjct: 302 HTKEEIGCSHPNCPLRAAIVIPLEMHGSIVGTLKMYFTNPNDLTFVERQLAEGLANIFSS 361

Query: 361 QLALGQAEEATRLLQDAEMKSLQAQVNPHFLFNALNTIYGLIRMDSEKARKLVQDFSKVI 420
           Q+ LG+AE  ++LL+DAE+KSLQAQV+PHF FN++N I  L+R++SEKAR+L+ + S
Sbjct: 362 QIELGEAETQSKLLKDAEIKSLQAQVSPHFFFNSINPISALVRINSEKARELLLELSYFF 421

Query: 421 RANLQRAKQNLIPLHDELEQVNAYLALEEARFPNMVAFNLDNQTNSDDNLMIPPFTLQVL 480
           RANLQ +KQ+ I L  EL QV AYL+LE+AR+P     N++ +    D +++PPF +Q+L
Sbjct: 422 RANLQGSKQHTITLDKELSQVRAYLSLEQARYPGRFNININVEDKYRD-VLVPPFLIQIL 480

Query: 481 IENSYKHAFKHVNKNNQLKVTIARNN-DRLHIIVQDNGIGIPKEKLITLGKKTQISKQGS 539
           +EN+ KHAF +  + N + V++ +     + IIVQDNG GI K+K+  LG+ +  S+ G+
Sbjct: 481 VENAIKHAFTNRKQGNDIDVSVIKETATHVRIIVQDNGQGISKDKMHLLGETSVESESGT 540

Query: 540 GTAIENLVRRLNIIYDGQASLKFESNDSGT 569
           G+A+ENL  RL  ++   A+L+FES  SGT
Sbjct: 541 GSALENLNLRLKGLFGKSAALQFESTSSGT 570
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1963> which encodes the amino acid sequence <SEQ ID 1964>. Analysis of this protein sequence reveals the following:

```
       Possible site: 39

  >>> Seems to have no N-terminal signal sequence
     INTEGRAL    Likelihood = -6.79   Transmembrane  283 - 299 ( 276 - 307)
     INTEGRAL    Likelihood = -5.57   Transmembrane   27 -  43 (  24 -  48)

  ----- Final Results -----
              bacterial membrane --- Certainty=0.3718(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:CAB54576 GB:AJ006396 histidine kinase [Streptococcus pneumoniae]
 Identities = 115/231 (49%), Positives = 159/231 (68%), Gaps = 7/231 (3%)

Query: 351 MLASIKAYIDEVYVLEVEQRDAQMRALQSQINPHFLYNTLEYIRMYALSCQQEELADVIY 410
           ML  ++  I ++Y LE+ Q+DA MRALQ+QINPHF+YNTLE++RMYA+   Q+ELAD+IY
Sbjct: 1   MLDRLEKNIHDIYQLELSQKDANMRALQAQINPHFMYNTLEFLRMYAVMQSQDELADIIY 60

Query: 411 AFASLLRNNISQDKMTTLKEELAFCEKYIYLYQMRYPDSFAYHVKIDESVADLAIPKFVI 470
            F+SLLRNNIS ++ T LK+EL FC KY YL  +RYP S AY  KID  + ++ IPKF +
Sbjct: 61  EFSSLLRNNISDERETLLKQELEFCRKYSYLCMVRYPKSIAYGFKIDPELENMKIPKFTL 120

Query: 471 QPLVENYFVHGIDYSRHDNALSIKALDETDHLLIQVLDNGRGISQERLADMEKRLQ---- 526
           QPLVENYF HG+D+ R DN +SIKAL +   + I V+DNGRG+S E+LA++ ++L
Sbjct: 121 QPLVENYFAHGVDHRRTDNVISIKALKQDGFVEILVVDNGRGMSAEKLANIREKLSQRYF 180

Query: 527 EHQTT---GNSSIGLQNVYLRLFHHFRDRVSWSMAKEPNGGFIIQIRIRKD 574
           EHQ +      SIG+ NV+ R   +F DR + ++      G   +I I+ +
Sbjct: 181 EHQASYSDQRQSIGIVNVHERFVLYFGDRYAITIESAEQAGVQYRITIQDE 231
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 59/180 (32%), Positives = 97/180 (53%), Gaps = 8/180 (4%)

Query: 375 QDAEMKSLQAQVNPHFLFNALNTI--YGLIRMDSEKARKLVQDFSKVIRANLQRAKQNLI 432
           +DA+M++LQ+Q+NPHFL+N L  I  Y L     E A  ++  F+ ++R N+ + K   +
Sbjct: 370 RDAQMRALQSQINPHFLYNTLEYIRMYALSCQQEELA-DVIYAFASLLRNNISQDK--MT 426

Query: 433 PLHDELEQVNAYLALEEARFPNMVAFNLDNQTNSDDNLMIPPFTLQVLIENSYKHAFKHV 492
            L +EL     Y+ L + R+P+  A+++    +  D L IP F +Q L+EN + H    +
Sbjct: 427 TLKEELAFCEKYIYLYQMRYPDSFAYHVKIDESVAD-LAIPKFVIQPLVENYFVHGIDYS 485

Query: 493 NKNNQLKVTIARNNDRLHIIVQDNGIGIPKEKLITLGKKTQISKQ--GSGTAIENLVRRL 550
             +N L +      D L I V DNG GI +E+L  + K+ Q  +    S   ++N+  RL
Sbjct: 486 RHDNALSIKALDETDHLLIQVLDNGRGISQERLADMEKRLQEHQTTGNSSIGLQNVYLRL 545
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 632

A DNA sequence (GBSx0672) was identified in *S.agalactiae* <SEQ ID 1965> which encodes the amino acid sequence <SEQ ID 1966>. Analysis of this protein sequence reveals the following:

```
Possible site: 24
>>> May be a lipoprotein

----- Final Results -----
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9827> which encodes amino acid sequence <SEQ ID 9828> was also identified.

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 633

A DNA sequence (GBSx0673) was identified in *S.agalactiae* <SEQ ID 1967> which encodes the amino acid sequence <SEQ ID 1968>. Analysis of this protein sequence reveals the following:

```
Possible site: 57
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -9.55   Transmembrane   52 -  68 (  45 -  74)
   INTEGRAL    Likelihood = -9.18   Transmembrane   83 -  99 (  76 - 106)
   INTEGRAL    Likelihood = -8.76   Transmembrane  126 - 142 ( 118 - 146)
   INTEGRAL    Likelihood = -7.48   Transmembrane  174 - 190 ( 170 - 191)
   INTEGRAL    Likelihood = -3.66   Transmembrane  195 - 211 ( 193 - 212)
   INTEGRAL    Likelihood = -1.28   Transmembrane   24 -  40 (  24 -  40)

----- Final Results -----
              bacterial membrane --- Certainty=0.4821(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

A related GBS gene <SEQ ID 8625> and protein <SEQ ID 8626> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 8
McG: Discrim Score:     -8.54
GvH: Signal Score (-7.5): -5.6
     Possible site: 57
>>> Seems to have no N-terminal signal sequence
ALOM program   count: 6 value:  -9.55 threshold:  0.0
   INTEGRAL    Likelihood = -9.55   Transmembrane   52 -  68 (  45 -  74)
   INTEGRAL    Likelihood = -9.18   Transmembrane   83 -  99 (  76 - 106)
   INTEGRAL    Likelihood = -8.76   Transmembrane  126 - 142 ( 118 - 146)
   INTEGRAL    Likelihood = -7.48   Transmembrane  174 - 190 ( 170 - 191)
   INTEGRAL    Likelihood = -3.66   Transmembrane  195 - 211 ( 193 - 212)
   INTEGRAL    Likelihood = -1.28   Transmembrane   24 -  40 (  24 -  40)
   PERIPHERAL  Likelihood = 13.05     100
 modified ALOM score:   2.41

*** Reasoning Step: 3

----- Final Results -----
              bacterial membrane --- Certainty=0.4821(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 634

A DNA sequence (GBSx0674) was identified in *S.agalactiae* <SEQ ID 1969> which encodes the amino acid sequence <SEQ ID 1970>. Analysis of this protein sequence reveals the following:

```
Possible site: 51
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.53   Transmembrane   83 -  99 (  83 -  99)

----- Final Results -----
              bacterial membrane --- Certainty=0.1213(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 635

A DNA sequence (GBSx0675) was identified in *S.agalactiae* <SEQ ID 1971> which encodes the amino acid sequence <SEQ ID 1972>. Analysis of this protein sequence reveals the following:

```
Possible site: 23
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1902(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 636

A DNA sequence (GBSx0676) was identified in *S.agalactiae* <SEQ ID 1973> which encodes the amino acid sequence <SEQ ID 1974>. Analysis of this protein sequence reveals the following:

```
Possible site: 20
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.4763(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 637

A DNA sequence (GBSx0677) was identified in *S.agalactiae* <SEQ ID 1975> which encodes the amino acid sequence <SEQ ID 1976>. Analysis of this protein sequence reveals the following:

```
Possible site: 20
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.5089(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 638

A DNA sequence (GBSx0678) was identified in *S.agalactiae* <SEQ ID 1977> which encodes the amino acid sequence <SEQ ID 1978>. Analysis of this protein sequence reveals the following:

```
Possible site: 25
>>> May be a lipoprotein

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 1978 (GBS184) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 26 (lane 7; MW 21kDa), in Figure 168 (lane 14-16; MW 36kDa – thioredoxin fusion) and in Figure 238 (lane 9; MW 36kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 37 (lane 7; MW 46.4kDa).

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 639

A DNA sequence (GBSx0679) was identified in *S.agalactiae* <SEQ ID 1979> which encodes the amino acid sequence <SEQ ID 1980>. Analysis of this protein sequence reveals the following:

```
Possible site: 52
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2179(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 640

A DNA sequence (GBSx0680) was identified in *S.agalactiae* <SEQ ID 1981> which encodes the amino acid sequence <SEQ ID 1982>. This protein is predicted to be immunogenic secreted protein precursor. Analysis of this protein sequence reveals the following:

```
Possible site: 34
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2166(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 9351> which encodes amino acid sequence <SEQ ID 9352> was also identified.

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1983> which encodes the amino acid sequence <SEQ ID 1984>. Analysis of this protein sequence reveals the following:

```
     Possible site: 19
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood = -3.77   Transmembrane    9 -  25 (   5 -  27)

----- Final Results -----
               bacterial membrane --- Certainty=0.2508(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 64/86 (74%), Positives = 76/86 (87%)

Query: 1   MGNGGDWKNKPGYQTTHEAKTGYAISFSPGQAGADRTYGHVAIVEDVKEDGSIPISESNV 60
           MGNGGDW+ KPG+ TTH+ K GY +SF+PGQAGAD TYGHVA+VE +KEDGSI ISESNV
Sbjct: 452 MGNGGDWQRKPGFVTTHKPKVGYVVSFAPGQAGADATYGHVAVVEQIKEDGSILISESNV 511

Query: 61  LGLGTISYRTFSAAEAAQLTYVVGEK 86
           +GLGTISYRTF+A +A+ LTYVVG+K
Sbjct: 512 MGLGTISYRTFTAEQASLLTYVVGDK 537
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 641

A DNA sequence (GBSx0681) was identified in *S.agalactiae* <SEQ ID 1985> which encodes the amino acid sequence <SEQ ID 1986>. This protein is predicted to be immunogenic secreted protein precursor. Analysis of this protein sequence reveals the following:

```
Possible site: 40
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2495(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
AAB52379 GB:U31811 immunogenic secreted protein precursor [Streptococcus pyogenes]
 Identities = 133/259 (51%), Positives = 170/259 (65%), Gaps = 4/259 (1%)

Query: 3   PSQPQVTATPQKSEVVTPAITSGIDLPDVAIPTAMASAAYVKHWIGNDAYTHNLLSHRYG 62
           P QP + A    + V  P   S  DL  +  P++ +SAAYV+HW G+ AYTHNLLS RYG
Sbjct: 174 PIQPPLGAA---APVFAPWRESDKDLSKLK-PSSRSSAAYVRHWTGDSAYTHNLLSRRYG 229

Query: 63  ITAAQLDGFLQSTGITYDSSRIDGQKILDREKSSGLDARAIIAIAIAESSLGTQGVATAP 122
           ITA QLDGFL S GI YD  R++G+++L+ EK +GLD RAI+AIA+AESSLGTQGVA
```

```
Sbjct: 230 ITAEQLDGFLNSLGIHYDKERLNGKRLLEWEKLTGLDVRAIVAIAMAESSLGTQGVAKEK 289

Query: 123 GANMFGFGAVDNNTTNAQNFSDDKAVIKMTQETIIQNQNTSFAIQDQKAQFLSTGNLNVA 182
           G+NMFG+GA D N  NA+ +SD+ A+  M ++TII N+N +F  QD KA+  S G L+
Sbjct: 290 GSNMFGYGAFDFNPNNAKKYSDEVAIRHMVEDTIIANKNQTFERQDLKAKKWSLGQLDTL 349

Query: 183 ARGGVYFTDASGSGKRRAAIMESIDKWIDAHGGISEISKELLNTSSVAMMAVPTSYSVSR 242
             GGVYFTD SGSG+RRA IM  +D+WID HG   +I + L  TS      VP  Y  S+
Sbjct: 350 IDGGVYFTDTSGSGQRRADIMTKLDQWIDDHGNTPDIPEHLKITSGTQFSEVPVGYKRSQ 409

Query: 243 ANQAGNYVAGTYPWGQRTW 261
                 Y + TY +GQ TW
Sbjct: 410 PQNVLTYKSETYSFGQCTW 428
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 1987> which encodes the amino acid sequence <SEQ ID 1988>. Analysis of this protein sequence reveals the following:

```
Possible site: 22
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
               bacterial outside --- Certainty=0.3000(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 143/265 (53%), Positives = 184/265 (68%), Gaps = 5/265 (1%)

Query: 2   VPSQPQVTATPQKSEVVTPA-----ITSGIDLPDVAIPTAMASAAYVKHWIGNDAYTHNL 56
           V + P  + + Q  E  TP        S +DL ++ IP+    AAYV+HW G +AYTH+L
Sbjct: 135 VDTAPASSLSKQLPEARTPIQSLSPYVSDLDLSEIDIPSVNTYAAYVEHWSGKNAYTHHL 194

Query: 57  LSHRYGITAAQLDGFLQSTGITYDSSRIDGQKILDREKSSGLDARAIIAIAIAESSLGTQ 116
           LS RYGI A Q+D +L+STGI YDS+RI+G+K+L  EK SGLD RAI+AIA++ESSLGTQ
Sbjct: 195 LSRRYGIKADQIDSYLKSTGIAYDSTRINGEKLLQWEKKSGLDVRAIVAIAMSESSLGTQ 254

Query: 117 GVATAPGANMFGFGAVDNNTTNAQNFSDDKAVIKMTQETIIQNQNTSFAIQDQKAQFLST 176
           G+AT  GANMFG+ A D + T A  F+DD A++KMTQ+TII+N+N++FA+QD KA   S
Sbjct: 255 GIATLLGANMFGYAAFDLDPTQASKFNDDSAIVKMTQDTIIKNKNSNFALQDLKAAKFSR 314

Query: 177 GNLNVAARGGVYFTDASGSGKRRAAIMESIDKWIDAHGGISEISKELLNTSSVAMMAVPT 236
           G LN A+ GGVYFTD +GSGKRRA IME +DKWID HGG   I  EL   SS +  +VP
Sbjct: 315 GQLNFASDGGVYFTDTTGSGKRRAQIMEDLDKWIDDHGGTPAIPAELKVQSSASFASVPA 374

Query: 237 SYSVSRANQAGNYVAGTYPWGQRTW 261
            Y +S++     Y A +Y WGQ TW
Sbjct: 375 GYKLSKSYDVLGYQASSYAWGQCTW 399
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 642

A DNA sequence (GBSx0682) was identified in *S.agalactiae* <SEQ ID 1989> which encodes the amino acid sequence <SEQ ID 1990>. Analysis of this protein sequence reveals the following:

```
Possible site: 27
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
               bacterial outside --- Certainty=0.3000(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

A related GBS gene <SEQ ID 8627> and protein <SEQ ID 8628> were also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 4
McG: Discrim Score:     11.56
GvH: Signal Score (-7.5): 0.870001
     Possible site: 27
>>> Seems to have a cleavable N-term signal seq.
ALOM program   count: 0 value:  11.88 threshold:  0.0
   PERIPHERAL  Likelihood = 11.88      63
 modified ALOM score:  -2.88

*** Reasoning Step: 3

----- Final Results -----
               bacterial outside --- Certainty=0.3000(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

SEQ ID 8628 (GBS159) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 28 (lane 4; MW 26kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 34 (lane 2; MW 41kDa).

GBS159-GST was purified as shown in Figure 198, lane 9.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 643

A DNA sequence (GBSx0683) was identified in *S.agalactiae* <SEQ ID 1991> which encodes the amino acid sequence <SEQ ID 1992>. Analysis of this protein sequence reveals the following:

```
Possible site: 32
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
             bacterial cytoplasm --- Certainty=0.2668(Affirmative) < succ>
              bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAB04699 GB:AP001510 unknown conserved protein [Bacillus halodurans]
 Identities = 32/76 (42%), Positives = 54/76 (70%)

Query: 7  LGSVIELKNDSQKVMITSRFPLYDNEGQLGYFDYSGCIFPISIVGNETYFFNLEDIDKVL 66
          +GS++ LK  + K+MI +R P+ +  G+   FDYSGC +P  +V ++ ++FN E+ID+V+
Sbjct: 4  IGSIVYLKEGTSKLMILNRGPILEANGENKMFDYSGCFYPQGLVPDKVFYFNHENIDEVV 63

Query: 67 FEGYYDENEEEMQKIF 82
          FEG+ D+ E+  QK+F
Sbjct: 64 FEGFQDDEEQRFQKLF 79
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 644

A DNA sequence (GBSx0684) was identified in *S.agalactiae* <SEQ ID 1993> which encodes the amino acid sequence <SEQ ID 1994>. Analysis of this protein sequence reveals the following:

```
Possible site: 32
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood =-14.81   Transmembrane   75 -  91 (  69 -  99)
   INTEGRAL    Likelihood =-14.38   Transmembrane  134 - 150 ( 129 - 179)
   INTEGRAL    Likelihood = -8.49   Transmembrane  157 - 173 ( 151 - 179)
   INTEGRAL    Likelihood = -1.17   Transmembrane   50 -  66 (  46 -  67)

----- Final Results -----
               bacterial membrane --- Certainty=0.6922(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 645

A DNA sequence (GBSx0685) was identified in *S.agalactiae* <SEQ ID 1995> which encodes the amino acid sequence <SEQ ID 1996>. Analysis of this protein sequence reveals the following:

```
Possible site: 35
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -0.11   Transmembrane   40 -  56 (  40 -  56)

----- Final Results -----
               bacterial membrane --- Certainty=0.1044(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 1996 (GBS204) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 49 (lane 13; MW 32kDa) and Figure 53 (lane 2; MW 14.7kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 54 (lane 6; MW 39.7kDa).

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 646

A DNA sequence (GBSx0686) was identified in *S.agalactiae* <SEQ ID 1997> which encodes the amino acid sequence <SEQ ID 1998>. Analysis of this protein sequence reveals the following:

```
Possible site: 38
>>> Seems to have no N-terminal signal sequence (or signal = aa 1-26)

----- Final Results -----
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

```
                  bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAC16670 GB:AJ302698 hypothetical protein [Staphylococcus
           haemolyticus]
 Identities = 60/254 (23%), Positives = 109/254 (42%), Gaps = 14/254 (5%)

Query: 2   VKVSVSSVGTQASTVAISMFSRVSALNDAITKLSSFAEAATLQGTAYSNAKSYATGTLTP 61
           + + V    +Q+S V  ++ S  S  +     +  F  A+ LQG AY + K + +  + P
Sbjct: 3   IDMYVGKSKSQSSDVGSTVKSISSGYDSLQKGIMQFVGASELQGQAYDSGKQFFSAVIAP 62

Query: 62  MLQGMILFSETLSEKCTELQTLYVSICGDEDLDSVVLESKLASDRASLKIAEALLEHLND 121
           + + +    E   + C +    Y S    + L    L   +      +   EA+   L
Sbjct: 63  LTESIKTLGELTEQACNDFVDQYQSEVDSQSLKESELLEDIEELNKQISQLEAMNASLKH 122

Query: 122 DPEPSKSAISSTKSNIKKLKKRIKSNQKKLDNLNEFNAHSATVFADISNAQSTVNQALAA 181
               + S +S     I  L+++ K  ++KL  L +F+A S  +F ++ + Q TV Q +
Sbjct: 123 KSSKNSSLLSGNHQMISSLEQQKKELEEKLRKLRQFDAKSPNIFKEVESFQKTVQQGINQ 182

Query: 182 VSTGFSGYNSKTGAFGKPTSGQMEWTKTVKKNWKEREDAKAEELKSKKAEESKKASKIEN 241
             T    ++     F  P    MEW K  ++   E    K +++ ++KA++ KK SK +
Sbjct: 183 AKT---AWDPGKQTFNIPAGKDMEWAKVSQQKALE---VKMDKI-NQKAKDGKKLSKNDI 235

Query: 242 TT-------KKSNV 248
            T       KKSN+
Sbjct: 236 FTIIAYQQQKKSNI 249
```

No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 1998 (GBS270) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 51 (lane 2; MW 34.3kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 54 (lane 7; MW 59.2kDa).

The GBS270-GST fusion product was purified (Figure 206, lane 3) and used to immunise mice. The resulting antiserum was used for FACS (Figure 265), which confirmed that the protein is immunoaccessible on GBS bacteria.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 647

A DNA sequence (GBSx0687) was identified in *S.agalactiae* <SEQ ID 1999> which encodes the amino acid sequence <SEQ ID 2000>. This protein is predicted to be outer surface protein F. Analysis of this protein sequence reveals the following:

```
Possible site: 23
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3323(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 2000 (GBS316) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 51 (lane 3; MW 23kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 55 (lane 2; MW 41.8kDa).

GBS316-GST was purified as shown in Figure 206, lane 4.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 648

A DNA sequence (GBSx0688) was identified in *S.agalactiae* <SEQ ID 2001> which encodes the amino acid sequence <SEQ ID 2002>. This protein is predicted to be actin-like protein arp3 (act4). Analysis of this protein sequence reveals the following:

```
Possible site: 17
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.0217(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 649

A DNA sequence (GBSx0689) was identified in *S.agalactiae* <SEQ ID 2003> which encodes the amino acid sequence <SEQ ID 2004>. This protein is predicted to be diarrheal toxin. Analysis of this protein sequence reveals the following:

```
Possible site: 25
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -8.65   Transmembrane   65 -  81 (  61 -  84)
   INTEGRAL    Likelihood = -3.98   Transmembrane   89 - 105 (  85 - 106)

----- Final Results -----
               bacterial membrane --- Certainty=0.4461(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB15175 GB:Z99120 alternate gene name: yueA~similar to
            hypothetical proteins [Bacillus subtilis]
 Identities = 452/1058 (42%), Positives = 664/1058 (62%), Gaps = 39/1058 (3%)

Query: 98    VTMIFSITGYFKNRKQYKQDLQERIDSYHDYLSDKSIELQKLAKEQKRGQHYHYPTIEGL 157
             +T+I S   YF+++ Q K+  ++R   Y  YL +K  ELQ LA++QK+   +H+P+ E +
Sbjct: 1     MTLITSTVQYFRDKNQRKKREEKRERVYKLYLDNKRKELQALAEKQKQVLEFHFPSFEQM 60

Query: 158   QEMADTYHHRIYEKTPLHFDFLYYRLGLGEVPTSYNIHYSQPERSGKK-DPLENEGYNLY 216
             + +      RI+EK+    D+L  RLG G VP+SY I+ S  + + +  D L  +  ++
Sbjct: 61    KYLTSEISDRIWEKSLESKDYLQLRLGTGTVPSSYEINMSGGDLANRDIDDLMEKSQHMQ 120

Query: 217   FNNRYIKNMPIVANLSHGPVGYIGPRGLVLEQLQLMVNQLAFFHSYHDVQFITIVPEEEM 276
```

```
               + I+N P+  +L+ GP+G +G   +V  ++  ++ QL+FF+SYHD++F+ I  EEE
Sbjct: 121  RVYKDIRNAPVTVDLAEGPMGLVGKSQIVKNEIHQLIGQLSFFNSYHDLRFVFIFHEEEY 180

Query: 277  DKWSWMRWLPHETLQDVNVRGFVYNQRSRDQVLNSLNQILKLRRTQREDKSAKEGTLFSP 336
              W WM+ +P   +  +  +GF+YN+++RDQ+L+SL ++++    +R+ +  KE   F P
Sbjct: 181  KDWEWMKCVPQFQMPHIYAKGFIYNEQTRDQLLSSLYELIR----ERDLEDDKEKLQFKP 236

Query: 337  HYVVIVTDEKLILDHVIMEFFTEDPTELGCSLIFVQDVMSSLSENIKTIINIKDRNTGQL 396
            H+V ++T+++LI +HVI+E+       LG S I   +   SLSENI T++   + + G +
Sbjct: 237  HFVFVITNQQLISEHVILEYLEGQHEHLGISTIVAAETKESLSENITTLVRYINEHEGDI 296

Query: 397  VIEEGELKETDFELDHFLEDYDKENISRRLAPLNHLQNLKSSIPEAVTFMEMYQAEEFED 456
            +I++ +     F LDH   + D E  SR L  LNH   + +SIPE V+F+E++ A+E ++
Sbjct: 297  LIQKKKAVRIPFRLDHHQRE-DNERFSRTLRTLNHQVGITNSIPETVSFLELFHAKEVKE 355

Query: 457  LHVQERWISHAPYKSSAVPLGLRGQDDIVYLNLHEKAHGPHGLVAGTTGSGKSEIIQSYI 516
            + +Q+RW++    KS +VP+G +G+DDIVYLNLHEKAHGPHGL+AGTTGSGKSE +Q+YI
Sbjct: 356  IGIQQRWLTSESSKSLSVPIGYKGKDDIVYLNLHEKAHGPHGLLAGTTGSGKSEFLQTYI 415

Query: 517  LSLAVNFHPHDVAFLLIDYKGGGMANLFKDLPHLLGTITNLDGAQ--SMRALVSINAELK 574
            LSLAV+FHPH+ AFLLIDYKGGGMA  F+++PHLLGTITN++G++  SMRAL SI +ELK
Sbjct: 416  LSLAVHFHPHEAAFLLIDYKGGGMAQPFRNIPHLLGTITNIEGSKNFSMRALASIKSELK 475

Query: 575  RRQRLFAKADVNHINQYQKKYKLGEVSEPMPHLFLISDEFAELKSNQPEFMKELVSTARI 634
            +RQRLF +  VNHIN Y K YK G+    MPHLFLISDEFAELKS +P+F++ELVS ARI
Sbjct: 476  KRQRLFDQYQVNHINDYTKLYKQGKAEVAMPHLFLISDEFAELKSEEPDFIRELVSAARI 535

Query: 635  GRSLGIHLILATQKPSGVVDDQIWSNSRFKLALKVADRGDSMEMLHTPDAAEITQAGRAY 694
            GRSLG+HLILATQKP G++DDQIWSNSRFK+ALKV D  DS E+L   DAA IT  GR Y
Sbjct: 536  GRSLGVHLILATQKPGGIIDDQIWSNSRFKVALKVQDATDSKEILKNSDAANITVTGRGY 595

Query: 695  LQVGNNEVYELFQSAWSGADYQPEKDDQGIEDHTIYSINDLGQYEILNDDLSGLDQAENI 754
            LQVGNNEVYELFQSAWSGA Y  E    G ED  I  + D G        LS +D  +N
Sbjct: 596  LQVGNNEVYELFQSAWSGAPYLEEV--YGTEDE-IAIVTDTGLI-----PLSEVDTEDNA 647

Query: 755  -KEVPTELDAIVENIQALTKEMGISDLPQPWLPPLSNQIAVTDLRKEGSVDLWSKAPSYK 813
             K+V TE++A+V+ I+ +  EMGI  LP PWLPPL+ +I  T         L+
Sbjct: 648  KKDVQTEIEAVVDEIERIQDEMGIEKLPSPWLPPLAERIPRT---------LFPSNEKDH 698

Query: 814  AVLGFMDIPSQQAQEVAYHDFEDDGHLSIFAGPSMGKSTALQTVTMDLARHNSPEFLNLY 873
                ++D P  Q Q    +   +DG++ IF     GKS A  T  M  A   +PE L++Y
Sbjct: 699  FHFAYVDEPDLQRQAPIAYKMMEDGNIGIFGSSGYGKSIAAATFLMSFADVYTPEELHVY 758

Query: 874  LFDFGTNGLLPLRRLPHVADFFTIDDDEKIAKFIARIKVEMSDRKKALSRYNVATAKLYR 933
            +FDFG   LLPL +LPH AD+F +D   KI KF+ RIK E+  RK+      ++  K+Y
Sbjct: 759  IFDFGNGTLLPLAKLPHTADYFLMDQSRKIEKFMIRIKEEIDRRKRLFREKEISHIKMYN 818

Query: 934  QVSGETMPQILIVIDSYEGLREAQTPTNLEACFQNISRDGSSLGISLVISAGRTAALRSS 993
             +S E +P I I ID+++ +++      LE+ F  +SRDG SLGI  +++A R  A+R S
Sbjct: 819  ALSEEELPFIFITIDNFDIVKDEM--HELESEFVQLSRDGQSLGIYFMLTATRVNAVRQS 876

Query: 994  LMANLKERIALKLTDDSESRTLVGRHQHIMEDIPGRGLIKRDDIEVLQVALSTEGTETFD 1053
            L+ NLK +I   L D SE  ++ GR +  +E IPGR +I+++++   Q+ L  +  +
Sbjct: 877  LLNNLKTKIVHYLMDQSEGYSIYGRPKFNLEPIPGRVIIQKEELYFAQMFLPVDADDDIG 936

Query: 1054 IINNIQNESDAMNSKWTG-PRPKAIPIVPEELTFDDFMATDSVQADLSANRL--PLGLEM 1110
            + N ++++   +  ++    +P  IP++PE L+  +     S++  L    L  P+GL
Sbjct: 937  MFNELKSDVQKLQGRFASMEQPAPIPMLPESLSTREL----SIRFKLERKPLSVPIGLHE 992

Query: 1111 VDVESYSLALNRFKHMLYMSDSDESLEAVGSHIIKVLL 1148
              V      L + KH L +  +        ++++KV+L
Sbjct: 993  ETVSPVYFDLGKHKHCLILGQTQRG----KTNVLKVML 1026
```

There is also homology to SEQ ID 24.

A related GBS gene <SEQ ID 8629> and protein <SEQ ID 8630> were also identified. Analysis of this protein sequence reveals the following:

```
Homology to a bacterial toxin
```

The protein has homology with the following sequences in the databases:

```
>OMNI|NT01BS3725 diarrheal toxin

 Score =  203 bits (511), Expect = 4e-51
 Identities = 123/377 (32%), Positives = 198/377 (51%), Gaps = 22/377 (5%)

Query: 1    MGISDLPQPWLPPLSNQIAVTDLRKEGSVDLWSKAPSYKAVLGFMDIPSQQAQEVAYHDF 60
            MGI  LP PWLPPL+ +I  T         L+           ++D P  Q Q    +
Sbjct: 704  MGIEKLPSPWLPPLAERIPRT---------LFPSNEKDHFHFAYVDEPDLQRQAPIAYKM 754

Query: 61   EDDGHLSIFAGPSMGKSTALQTVTMDLARHNSPEFLNLYLFDFGTNGLLPLRRLPHVADF 120
             +DG++ IF     GKS A  T  M  A   +PE L++Y+FDFG   LLPL +LPH AD+
Sbjct: 755  MEDGNIGIFGSSGYGKSIAAATFLMSFADVYTPEELHVYIFDFGNGTLLPLAKLPHTADY 814

Query: 121  FTIDDDEKIAKFIARIKVEMSDRKKALSRYNVATAKLYRQVSGETMPQILIVIDSYEGLR 180
            F +D   KI KF+ RIK E+  RK+      ++  K+Y  +S E +P I I ID+++ ++
Sbjct: 815  FLMDQSRKIEKFMIRIKEEIDRRKRLFREKEISHIKMYNALSEEELPFIFITIDNFDIVK 874

Query: 181  EAQTPTNLEACFQNISRDGSSLGISLVISAGRTAALRSSLMANLKERIALKLTDDSESRT 240
            +      LE+ F  +SRDG SLGI  +++A R  A+R SL+ NLK +I   L D SE  +
Sbjct: 875  DEM--HELESEFVQLSRDGQSLGIYFMLTATRVNAVRQSLLNNLKTKIVHYLMDQSEGYS 932

Query: 241  LVGRHQHIMEDIPGRGLIKRDDIEVLQVALSTEGTETFDIINNIQNESDAMNSKWTG-PR 299
            + GR +  +E IPGR +I+++++   Q+ L  +  +   + N ++++   +  ++    +
Sbjct: 933  IYGRPKFNLEPIPGRVIIQKEELYFAQMFLPVDADDDIGMFNELKSDVQKLQGRFASMEQ 992

Query: 300  PKAIPIVPEELTFDDFMATDSVQADLSANRL--PLGLEMVDVESYSLALNRFKHMLYMSD 357
            P  IP++PE L+  +     S++  L    L  P+GL    V      L + KH L +
Sbjct: 993  PAPIPMLPESLSTREL----SIRFKLERKPLSVPIGLHEETVSPVYFDLGKHKHCLILGQ 1048

Query: 358  SDESLEAVGSHIIKVLL 374
            +        ++++KV+L
Sbjct: 1049 TQRG----KTNVLKVML 1061
```

SEQ ID 8630 (GBS326) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 65 (lane 5; MW 66kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 71 (lane 5; MW 91kDa).

GBS326-GST was purified as shown in Figure 212, lane 5.

GBS326LN was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 127 (lane 2-4; MW 114kDa). It was also expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 184 (lane 6; MW 114kDa). The purified protein is shown in Figure 236, lane 12.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 650

A DNA sequence (GBSx0690) was identified in *S.agalactiae* <SEQ ID 2005> which encodes the amino acid sequence <SEQ ID 2006>. Analysis of this protein sequence reveals the following:

```
Possible site: 16
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2693(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes.*

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 651

A DNA sequence (GBSx0691) was identified in *S.agalactiae* <SEQ ID 2007> which encodes the amino acid sequence <SEQ ID 2008>. Analysis of this protein sequence reveals the following:

```
Possible site: 38
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3933(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes.*

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 652

A DNA sequence (GBSx0692) was identified in *S.agalactiae* <SEQ ID 2009> which encodes the amino acid sequence <SEQ ID 2010>. Analysis of this protein sequence reveals the following:

```
Possible site: 55
>>> Seems to have no N-terminal signal sequence
   INTEGRAL    Likelihood = -7.32   Transmembrane  225 - 241 ( 219 - 246)

----- Final Results -----
               bacterial membrane --- Certainty=0.3930(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:BAB04693 GB:AP001510 unknown conserved protein [Bacillus halodurans]
 Identities = 83/320 (25%), Positives = 162/320 (49%), Gaps = 1/320 (0%)

Query: 103 VNFILHPSNLFLTKNATAKIAYRSLPGIMRPEKFGPEEFLYQFKCFVFALLTQHDYIELY 162
           ++ I+ P N+ ++        +  +   + P +  PE    + +  +  LL  +     Y
Sbjct: 106 LHLIVSPENVLVSDGLDVTFIHYGVKDSIPPYETDPERLFLELRATLLVLLDGNHRFHEY 165

Query: 163 NGAISVIEVSDFLKSIYHAETIQAVRDIITIDYEQQVEVETHTLAKVSRAKYKLYKYISV 222
                 +++S   KS+    T++ +R++I   + Q+ E +   L KV + K+ + K+  +
Sbjct: 166 MNYHDTLKLSPEAKSLVQQTTLEGLRELIR-HWIQEHEQQEKQLHKVPKTKWTIQKWAGI 224

Query: 223 WLGALSTILLIPLVYLVFIHNPFKEKMLAADTSFIKVDYNQVINRLEHVKVSKLPYTQKY 282
            L A     +I +VY++    P +E   A+  +++  +Y+QVI+ LE    + +P   KY
Sbjct: 225 GLIAALVPAIIYIVYVLAFLQPRQEAFTASHAAYLNENYSQVIDTLEPYSPNSMPRVVKY 284

Query: 283 ELAYSYINGMSFSEEQREVILNNVTLKTDELYLDYWINIGRGLDDDAIDAAKRLDDSDLV 342
           +LA SY+         RE + N + L+  E Y DYWI IGRG ++ AID A+ L D + +
Sbjct: 285 QLAQSYVAIEPLQAYHRENLKNVLVLQAAESYFDYWIAIGRGENEKAIDIARGLQDKEWL 344
```

```
Query: 343 IYAIVQKMDQVRKDNSLSGKDREQKLSELQTDYDKYWKDRKTALTDEESKSKNSNNHSTN 402
           +YA V++ ++V+ D +LSGK+RE  + E++ + D Y ++ +    + E+   N+   ++N
Sbjct: 345 VYANVKRREEVKSDENLSGKEREDLIKEIEAEIDDYMRELEELAEEGEAFQPNAEPAASN 404

Query: 403 SNKESSESSSTTASTSSKTK 422
             +E    +    S + + K
Sbjct: 405 ELEEDEGDTEEDDSDNQEAK 424
```

No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 2010 (GBS337) was expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 62 (lane 3; MW 50.3kDa).

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 653

A DNA sequence (GBSx0693) was identified in *S.agalactiae* <SEQ ID 2011> which encodes the amino acid sequence <SEQ ID 2012>. Analysis of this protein sequence reveals the following:

```
Possible site: 27
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood =-14.01   Transmembrane  131 - 147 ( 122 - 153)

----- Final Results -----
              bacterial membrane --- Certainty=0.6604(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 8631> which encodes amino acid sequence <SEQ ID 8632> was also identified. Analysis of this protein sequence reveals the following:

```
Lipop: Possible site: -1   Crend: 8
McG: Discrim Score:      13.38
GvH: Signal Score (-7.5): -1.25
     Possible site: 23
>>> Seems to have a cleavable N-term signal seq.
ALOM program   count: 1 value: -14.01 threshold:  0.0
   INTEGRAL    Likelihood =-14.01   Transmembrane  127 - 143 ( 118 - 149)
   PERIPHERAL  Likelihood = 16.13     113
 modified ALOM score:   3.30

*** Reasoning Step: 3

----- Final Results -----
              bacterial membrane --- Certainty=0.6604(Affirmative) < succ>
               bacterial outside --- Certainty=0.0000(Not Clear) < succ>
             bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

SEQ ID 8632 (GBS140) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 32 (lane 3; MW 43kDa). It was also expressed in *E.coli* as a His-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 49 (lane 8; MW 18kDa).

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 654

A DNA sequence (GBSx0694) was identified in *S.agalactiae* <SEQ ID 2013> which encodes the amino acid sequence <SEQ ID 2014>. Analysis of this protein sequence reveals the following:

```
Possible site: 15
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.1486(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 655

A DNA sequence (GBSx0695) was identified in *S.agalactiae* <SEQ ID 2015> which encodes the amino acid sequence <SEQ ID 2016>. Analysis of this protein sequence reveals the following:

```
Possible site: 32
>>> Seems to have an uncleavable N-term signal seq
   INTEGRAL    Likelihood =-14.59   Transmembrane  984 -1000 ( 976 -1009)
   INTEGRAL    Likelihood = -9.71   Transmembrane   19 -  35 (  15 -  42)
   INTEGRAL    Likelihood = -9.50   Transmembrane  872 - 888 ( 865 - 890)
   INTEGRAL    Likelihood = -6.37   Transmembrane  927 - 943 ( 924 - 951)
   INTEGRAL    Likelihood = -4.19   Transmembrane  831 - 847 ( 828 - 847)
   INTEGRAL    Likelihood = -2.87   Transmembrane  899 - 915 ( 899 - 916)

----- Final Results -----
               bacterial membrane --- Certainty=0.6838(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

A related GBS nucleic acid sequence <SEQ ID 8633> which encodes amino acid sequence <SEQ ID 8634> was also identified. Analysis of this protein sequence reveals the following:

```
Lipop Possible site: -1   Crend: 6
SRCFLG: 0
McG: Length  of  UR:   20
     Peak Value of UR:   3.40
     Net Charge of CR: 3
McG: Discrim Score:     13.67
GvH: Signal Score (-7.5): -3.27
     Possible site: 21
>>> Seems to have an uncleavable N-term signal seq
Amino Acid Composition: calculated from 1
ALOM program   count: 6 value: -14.59 threshold:  0.0
   INTEGRAL    Likelihood =-14.59   Transmembrane  973 - 989 ( 965 - 998)
   INTEGRAL    Likelihood = -9.71   Transmembrane    8 -  24 (   4 -  31)
   INTEGRAL    Likelihood = -9.50   Transmembrane  861 - 877 ( 854 - 879)
   INTEGRAL    Likelihood = -6.37   Transmembrane  916 - 932 ( 913 - 940)
   INTEGRAL    Likelihood = -4.19   Transmembrane  820 - 836 ( 817 - 836)
   INTEGRAL    Likelihood = -2.87   Transmembrane  888 - 904 ( 888 - 905)
   PERIPHERAL  Likelihood =  3.82     936
 modified ALOM score:   3.42
icml HYPID: 7  CFP: 0.684

*** Reasoning Step: 3
```

```
----- Final Results -----
               bacterial membrane --- Certainty=0.6838(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAB86324 GB:AE000938 phage infection protein homolog
           [Methanothermobacter thermoautotrophicus]
 Identities = 96/454 (21%), Positives = 190/454 (41%), Gaps = 63/454 (13%)

Query: 1   MLKIKYILGRIMKR-NNFRILWYIIAVALFLVAIAGLNLKLQGDHAKENKTTQSATNTKL 59
           M K   I  + MK   N  ++ ++IAV + + A+  +   +Q      ++T+       +
Sbjct: 1   MRKALEIFWKDMKTVKNSPVVLFVIAVIICIPALYAV-FNIQATLDPYSRTSS------I 53

Query: 60  NIALVNEDQNVSNGKESYNLGASYIKSIERDNSQNWSVVSRGTAQNGLDKGDYQLMVIIP 119
            +A+VNED          N+GA ++  + ++ + +W  V R  A +GL KG Y  ++IIP
Sbjct: 54  EVAVVNEDMGADFNGTHLNVGAEFVSELRKNRNFDWQFVDRSDAMDGLRKGKYYAVLIIP 113

Query: 120 NNFSQKLLDVNKANAEQTTISYKVNAKGNLALEKKATEKEKDIVSELNSHLVNMYMASIL 179
            NFS  LL +      Q +I Y VN K N    +        + +++NS +V      +
Sbjct: 114 GNFSSDLLSIKNGTPRQASIKYMVNDKLNPVAPRITNAGADALQAKINSEVVKTIDGIVF 173

Query: 180 SNLYTAQENVQA----------MVNVQSGNISNYQKNLLDSATNF---QNIFPAL----- 221
             +  A E  +A           VN  +GN+    + L  + ++    QN++ +L
Sbjct: 174 GKISEAGELARANRDDILRTKRFVNELNGNLGKIDETLSTANSDLEKGQNLWSSLKTDLP 233

Query: 222 -VNQSSSSITANESLKKS------------LEASDNMFNDLVTTQTNTGKDLSSL----- 263
            +  +++ +    SL +S            +++ ++  ++ +T+       L+SL
Sbjct: 234 EIRDNANFVKEKYSLLESYIGKDPAKALSTVQSMESHLSEAITSMKYLRAVLASLYSATG 293

Query: 264 -------IEQRHQDSISYEAFSTSLLEMNNELLEKQLSDIITQAQKDQETLSSQLNSIMG 316
                  I+Q   +     +    L  + ++L  K  +D I + +   + + S LN +M
Sbjct: 294 DPKLKTAIDQIDTNIEKASSVLGILQTIESDLKTKGTTDRIVKLKASIDRMDSALNKLMD 353

Query: 317 D-DNNHNHKENSSAYLNVARQKIQELSEALKSQDNIAKDQSEQLDKIVREGLASYFAKNN 375
             D      +++SA L +A  +   +  A+       +D S +L+ I  + L S     +
Sbjct: 354 SRDEIDAAMQDASAKLGIANARWPTMRSAI-------QDASRKLNMISDDDLNSLVKLAD 406

Query: 376 KDNITLLELLKSHSTNEK----TLKDFKAKVADF 405
            D   + E  +S    EK     +K++ + +A F
Sbjct: 407 IDPSAVREYFRSPVRMEKEHIYPVKNYGSALAPF 440
```

SEQ ID 8634 (GBS250) was expressed in *E.coli* as a GST-fusion product. SDS-PAGE analysis of total cell extract is shown in Figure 47 (lane 4; MW 136kDa).

GBS250-GST was purified as shown in Figure 203, lane 4.

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 656

A DNA sequence (GBSx0696) was identified in *S.agalactiae* <SEQ ID 2019> which encodes the amino acid sequence <SEQ ID 2020>. Analysis of this protein sequence reveals the following:

```
Possible site: 39
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.5009(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAA46375 GB:X65276 ORFA1 [Clostridium acetobutylicum]
 Identities = 35/91 (38%), Positives = 53/91 (57%)

Query: 1  MAQIKLTPEELRSSAQKYTAGSQQVTEVLNLLTQEQAVIDENWDGSTFDSFEAQFNELSP 60
           MAQI +TPEEL+S AQ Y    +++ + +  +     + I E W G  F ++  Q+N+L
Sbjct: 1  MAQISVTPEELKSQAQVYIQSKEEIDQAIQKVNSMNSTIAEEWKGQAFQAYLEQYNQLHQ 60

Query: 61 KITEFAQLLEDINQQLLKVADIIEQTDADIA 91
            + +F  LLE +NQQL K AD + + DA  A
Sbjct: 61 TVVQFENLLESVNQQLNKYADTVAERDAQDA 91
```

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 657

A DNA sequence (GBSx0697) was identified in *S.agalactiae* <SEQ ID 2021> which encodes the amino acid sequence <SEQ ID 2022>. Analysis of this protein sequence reveals the following:

```
Possible site: 22
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3741(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has no significant homology with any sequences in the GENPEPT database.

No corresponding DNA sequence was identified in *S.pyogenes*.

Based on this analysis, it was predicted that this protein and its epitopes, could be useful antigens for vaccines or diagnostics.

Example 658

A repeated DNA sequence (GBSx0698) was identified in *S.agalactiae* <SEQ ID 2023> which encodes the amino acid sequence <SEQ ID 2024>. This protein is predicted to be carbamoylphosphate synthetase (carB). Analysis of this protein sequence reveals the following:

```
Possible site: 23
>>> Seems to have a cleavable N-term signal seq.
   INTEGRAL    Likelihood = -1.33   Transmembrane  807 - 823 ( 807 - 823)

----- Final Results -----
               bacterial membrane --- Certainty=0.1532(Affirmative) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAA03928 GB:AJ000109 carbamoylphosphate synthetase [Lactococcus
            lactis]
 Identities = 771/1062 (72%), Positives = 901/1062 (84%), Gaps = 5/1062 (0%)

Query: 1     MPKRTDIRKIMVIGSGPIVIGQAAEFDYSGTQACLSLKEEGYQVVLVNSNPATIMTDKDI 60
              MPKR DI+KIM+IGSGPI+IGQAAEFDY+GT+ACL+LKEEGY+VVLVNSNPATIMTD++I
Sbjct: 1     MPKRNDIKKIMIIGSGPIIIGQAAEFDYAGTEACLALKEEGYEVVLVNSNPATIMTDREI 60
```

```
Query: 61   ADKVYIEPITLEFVTRILRKERPDALLPTLGGQTGLNMAMALSKNGILEELNVELLGTKL 120
            AD VYIEPITLEFV++ILRKERPDALLPTLGGQTGLNMAM LSK GILEELNVELLGTKL
Sbjct: 61   ADTVYIEPITLEFVSKILRKERPDALLPTLGGQTGLNMAMELSKTGILEELNVELLGTKL 120

Query: 121  SAIDKAEDRDLFKQLMEELNQPIPESEIVNSVEEAIQFAEQIGYPLIVRPAFTLGGTGGG 180
            SAID+AEDR+LFK+L E +N+P+  S+I  +VEEAI  A++IGYP+IV PAFT+GGTGGG
Sbjct: 121  SAIDQAEDRELFKELCESINEPLCASDIATTVEEAINIADKIGYPIIVGPAFTMGGTGGG 180

Query: 181  MCDNQEQLVDITTKGLKLSPVTQCLIERSIAGFKEIEYEVMRDAADNALVVCNMENFDPV 240
            +CD +E+L +I   GLKLSPVTQCLIE SIAG+KEIEYEVMRD+ADNA+VVCNMENFDPV
Sbjct: 181  ICDTEEELREIVANGLKLSPVTQCLIEESIAGYKEIEYEVMRDSADNAIVVCNMENFDPV 240

Query: 241  GIHTGDSIVFAPAQTLSDVENQLLRDASLDIIRALKIEGGCNVQLALDPNSFKYYVIEVN 300
            G+HTGDSIVFAP+QTLSD E Q+LRDASL+IIRALKIEGGCNVQLALDPNS++Y VIEVN
Sbjct: 241  GVHTGDSIVFAPSQTLSDNEYQMLRDASLNIIRALKIEGGCNVQLALDPNSYEYRVIEVN 300

Query: 301  PRVSRSSALASKATGYPIAKLAAKIAVGLTLDEVINPITKTTYAMFEPALDYVVAKMPRF 360
            PRVSRSSALASKATGYPIAK++AKIA+G+TLDE+INP+T  TYAMFEPALDYVVAK+ RF
Sbjct: 301  PRVSRSSALASKATGYPIAKMSAKIAIGMTLDEIINPVTNKTYAMFEPALDYVVAKIARF 360

Query: 361  PFDKFESGDRKLGTQMKATGEVMAIGRNIEESLLKACRSLEIGVDHIKIADLDNVSDDVL 420
            PFDKFE+GDR LGTQMKATGEVMAIGRNIEESLLKA RSLEIGV H ++ +     D+ L
Sbjct: 361  PFDKFENGDRHLGTQMKATGEVMAIGRNIEESLLKAVRSLEIGVFHNEMTEAIEADDEKL 420

Query: 421  LEKIRKAEDDRLFYLAEALRRHYSIEKLASLTSIDSFFLDKLRVIVELEDLLSKNRLDIN 480
             EK+ K +DDRLFY++EA+RR   IE++A LT ID FFLDKL  IVE+E+ L  N  +
Sbjct: 421  YEKMVKTQDDRLFYVSEAIRRGIPIEEIADLTKIDIFFLDKLLYIVEIENQLKVNIFEPE 480

Query: 481  ILKKVKNKGFSDKAIASLWQINEDQVRNMRKEAGILPVYKMVDTCASEFDSATPYFYSTY 540
            +LK  K  GFSD+ IA LW +  ++VR  R+E  I+PVYKMVDTCA+EF+S+TPYFYSTY
Sbjct: 481  LLKTAKKNGFSDREIAKLWNVTPEEVRRRRQENKIIPVYKMVDTCAAEFESSTPYFYSTY 540

Query: 541  AVENESLISDKASILVLGSGPIRIGQGVEFDYATVHSVKAIRESGFEAIIMNSNPETVST 600
              ENES  SDK  I+VLGSGPIRIGQGVEFDYATVH VKAI+  G EAI++NSNPETVST
Sbjct: 541  EWENESKRSDKEKIIVLGSGPIRIGQGVEFDYATVHCVKAIQALGKEAIVINSNPETVST 600

Query: 601  DFSISDKLYFEPLTFEDVMNVIDLEKPEGVILQFGGQTAINLAKDLNKAGVKILGTQLED 660
            DFSISDKLYFEPLTFEDVMNVIDLE+P  VI+QFGGQTAINLA+ L+KAGVKILGTQ+ED
Sbjct: 601  DFSISDKLYFEPLTFEDVMNVIDLEEPLVVIVQFGGQTAINLAEHLSKAGVKILGTQVED 660

Query: 661  LDRAENRKQFEATLQALNIPQPPGFTATTEEEAVNAAQKIGYPVLVRPSYVLGGRAMKIV 720
            LDRAE+R  FE  LQ L+IPQPPG TAT EEEAV  A KIGYPVL+RPS+VLGGRAM+I+
Sbjct: 661  LDRAEDRDLFEKALQDLDIPQPPGATATNEEEAVANANKIGYPVLIRPSFVLGGRAMEII 720

Query: 721  ENEEDLRHYMTTAVKASPDHPVLIDAYLIGKECEVDAISDGQNILIPGIMEHIERSGVHS 780
             NE+DLR YM  AVKASP+HPVL+D+YL G+ECEVDAI DG+ +L+PGIMEHIER+GVHS
Sbjct: 721  NNEKDLRDYMNRAVKASPEHPVLVDSYLQGQECEVDAICDGKEVLLPGIMEHIERAGVHS 780

Query: 781  GDSMAVYPPQTLSETIIETIVDYTKRLAIGLNCIGMMNIQFVIKDQKVYVIEVNPRASRT 840
            GDSMAVYPPQ LS+ II+TIVDYTKRLAIGLNCIGMMNIQFVI +++VYVIEVNPRASRT
Sbjct: 781  GDSMAVYPPQNLSQAIIDTIVDYTKRLAIGLNCIGMMNIQFVIYEEQVYVIEVNPRASRT 840

Query: 841  LPFLSKVTHIPMAQVATKVILGDKLCNFTYGYDLYPASDMVHIKAPVFSFTKLAKVDSLL 900
            +PFLSKVT+IPMAQ+AT++ILG+ L +  Y   L P  DMVH+KAPVFSFTKLAKVDSLL
Sbjct: 841  VPFLSKVTNIPMAQLATQMILGENLKDLGYEAGLAPTPDMVHVKAPVFSFTKLAKVDSLL 900

Query: 901  GPEMKSTGEVMGSDINLQKALYKAFEAAYLHMPDYGNIVFTVDDTDKEEALELAKVYQSI 960
            GPEMKSTG  MGSD+ L+KALYK+FEAA LHM DYG+++FTV D DKEE L LAK +  I
Sbjct: 901  GPEMKSTGLAMGSDVTLEKALYKSFEAAKLHMADYGSVLFTVADEDKEETLALAKDFAEI 960

Query: 961  GYRIYATQGTAIYFDANGLETVLVGKL--GENDRNHIPDLIKNGKIQAVINTVGQNNID- 1017
            GY + AT GTA +   NGL    V KL  GE++   + + I+ G++QAV+NT+G
Sbjct: 961  GYSLVATAGTAAFLKENGLYVREVEKLAGGEDEEGTLVEDIRQGRVQAVVNTMGNTRASL 1020

Query: 1018 --NHDALIIRRSAIEQGVPLFTSLDTAHAMFKVLESRAFTLK 1057
                D   IR+ AI +G+PLFTSLDT  A+ KV++SR+FT K
Sbjct: 1021 TTATDGFRIRQEAISRGIPLFTSLDTVAAILKVMQSRSFTTK 1062
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2025> which encodes the amino acid sequence <SEQ ID 2026>. Analysis of this protein sequence reveals the following:

```
        Possible site: 21
   >>> Seems to have no N-terminal signal sequence
      INTEGRAL    Likelihood = -1.17    Transmembrane  773 - 789 ( 773 - 789)

   ----- Final Results -----
                  bacterial membrane --- Certainty=0.1468(Affirmative) < succ>
                   bacterial outside --- Certainty=0.0000(Not Clear) < succ>
                 bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
>GP:CAA03928 GB:AJ000109 carbamoylphosphate synthetase [Lactococcus
            lactis]
 Identities = 753/1030 (73%), Positives = 876/1030 (84%), Gaps = 6/1030 (0%)

Query: 1     LALKEEGYKVILVNSNPATIMTDKEIADKVYIEPLTLEFVNRIIRKERPDAILPTLGGQT 60
             LALKEEGY+V+LVNSNPATIMTD+EIAD VYIEP+TLEFV++I+RKERPDA+LPTLGGQT
Sbjct: 35    LALKEEGYEVVLVNSNPATIMTDREIADTVYIEPITLEFVSKILRKERPDALLPTLGGQT 94

Query: 61    GLNMAMALSKAGILDDLEIELLGTKLSAIDQAEDRDLFKQLMQELDQPIPESTIVKTVDE 120
             GLNMAM LSK GIL++L +ELLGTKLSAIDQAEDR+LFK+L + +++P+  S I  TV+E
Sbjct: 95    GLNMAMELSKTGILEELNVELLGTKLSAIDQAEDRELFKELCESINEPLCASDIATTVEE 154

Query: 121   AVTFARDIGYPVIVRPAFTLGGTGGGICSSEEELCEITENGLKLSPVTQCLIERSIAGFK 180
             A+  A  IGYP+IV PAFT+GGTGGGIC +EEEL EI  NGLKLSPVTQCLIE SIAG+K
Sbjct: 155   AINIADKIGYPIIVGPAFTMGGTGGGICDTEEELREIVANGLKLSPVTQCLIEESIAGYK 214

Query: 181   EIEYEVMRDSADNALVVCNMENFDPVGIHTGDSIVFAPTQTLSDIENQMLRDASLKIIRA 240
             EIEYEVMRDSADNA+VVCNMENFDPVG+HTGDSIVFAP+QTLSD E QMLRDASL IIRA
Sbjct: 215   EIEYEVMRDSADNAIVVCNMENFDPVGVHTGDSIVFAPSQTLSDNEYQMLRDASLNIIRA 274

Query: 241   LKIEGGCNVQLALDPYSFKYYVIEVNPRVSRSSALASKATGYPIAKLAAKIAVGLTLDEM 300
             LKIEGGCNVQLALDP S++Y VIEVNPRVSRSSALASKATGYPIAK++AKIA+G+TLDE+
Sbjct: 275   LKIEGGCNVQLALDPNSYEYRVIEVNPRVSRSSALASKATGYPIAKMSAKIAIGMTLDEI 334

Query: 301   INPITGTTYAMFEPALDYVVAKIPRFPFDKFEHGERQLGTQMKATGEVMAIGRNLEESLL 360
             INP+T  TYAMFEPALDYVVAKI RFPFDKFE+G+R LGTQMKATGEVMAIGRN+EESLL
Sbjct: 335   INPVTNKTYAMFEPALDYVVAKIARFPFDKFENGDRHLGTQMKATGEVMAIGRNIEESLL 394

Query: 361   KACRSLEIGVCHNEMTSLSNISDEELVTKVIKAQDDRLFYLSEAIRRGYSIEELESLTKI 420
             KA RSLEIGV HNEMT      DE+L  K++K QDDRLFY+SEAIRRG  IEE+  LTKI
Sbjct: 395   KAVRSLEIGVFHNEMTEAIEADDEKLYEKMVKTQDDRLFYVSEAIRRGIPIEEIADLTKI 454

Query: 421   DLFFLDKLLHIVEIEQELQMHVDHLESLKKAKRYGFSDQKIAEIWQKDESDIRAMRHSHS 480
             D+FFLDKLL+IVEIE +L++++   E LK AK+ GFSD++IA++W     ++R  R  +
Sbjct: 455   DIFFLDKLLYIVEIENQLKVNIFEPELLKTAKKNGFSDREIAKLWNVTPEEVRRRRQENK 514

Query: 481   LYPVYKMVDTCAAEFDAKTPYFYSTYELENESVQSNKESILVLGSGPIRIGQGVEFDYAT 540
             + PVYKMVDTCAAEF++ TPYFYSTYE ENES +S+KE I+VLGSGPIRIGQGVEFDYAT
Sbjct: 515   IIPVYKMVDTCAAEFESSTPYFYSTYEWENESKRSDKEKIIVLGSGPIRIGQGVEFDYAT 574

Query: 541   VHSVKAIQKAGYEAIIMNSNPETVSTDFSVSDKLYFEPLTFEDVMNVIDLEQPKGVIVQF 600
             VH VKAIQ  G EAI++NSNPETVSTDFS+SDKLYFEPLTFEDVMNVIDLE+P  VIVQF
Sbjct: 575   VHCVKAIQALGKEAIVINSNPETVSTDFSISDKLYFEPLTFEDVMNVIDLEEPLVVIVQF 634

Query: 601   GGQTAINLAQALSEAGVTILGTQVEDLDRAEDRDLFEKALKELGIPQPQGQTATNEEEAL 660
             GGQTAINLA+ LS+AGV ILGTQVEDLDRAEDRDLFEKAL++L IPQP G TATNEEEA+
Sbjct: 635   GGQTAINLAEHLSKAGVKILGTQVEDLDRAEDRDLFEKALQDLDIPQPPGATATNEEEAV 694

Query: 661   EAAKKIGFPVLVRPSYVLGGRAMEIVENKEDLREYIRTAVKASPEHPILVDSYIFGKECE 720
               A KIG+PVL+RPS+VLGGRAMEI+ N++DLR+Y+  AVKASPEHP+LVDSY+ G+ECE
Sbjct: 695   ANANKIGYPVLIRPSFVLGGRAMEIINNEKDLRDYMNRAVKASPEHPVLVDSYLQGQECE 754

Query: 721   VDAISDGKSVLIPGIMEHIERAGVHSGDSMAVYPPQQLSKQIQETIAEYTKRLAIGLNCI 780
```

```
            VDAI DGK VL+PGIMEHIERAGVHSGDSMAVYPPQ LS+ I +TI +YTKRLAIGLNCI
Sbjct: 755  VDAICDGKEVLLPGIMEHIERAGVHSGDSMAVYPPQNLSQAIIDTIVDYTKRLAIGLNCI 814

Query: 781  GMMNVQFVIKNEQVYVIEVNPRASRTVPFLSKVTGIPMAQIATKLILGQTLKDLGYEDGL 840
            GMMN+QFVI  EQVYVIEVNPRASRTVPFLSKVT IPMAQ+AT++ILG+ LKDLGYE GL
Sbjct: 815  GMMNIQFVIYEEQVYVIEVNPRASRTVPFLSKVTNIPMAQLATQMILGENLKDLGYEAGL 874

Query: 841  YPQSPLVHIKAPVFSFTKLAQVDSLLGPEMKSTGEVMGSDTSLEKALYKAFEANNSHLSE 900
             P   +VH+KAPVFSFTKLA+VDSLLGPEMKSTG  MGSD +LEKALYK+FEA   H+++
Sbjct: 875  APTPDMVHVKAPVFSFTKLAKVDSLLGPEMKSTGLAMGSDVTLEKALYKSFEAAKLHMAD 934

Query: 901  FGQIVFTIADDSKAEALSLARRFKAIGYQIMATQGTAAYFAEQGLSACLVGKIGDAANDI 960
            +G ++FT+AD+ K E L+LA+ F  IGY ++AT GTAA+  E GL    V K+    ++
Sbjct: 935  YGSVLFTVADEDKEETLALAKDFAEIGYSLVATAGTAAFLKENGLYVREVEKLAGGEDEE 994

Query: 961  PTLV---RHGHVQAIVNTVGIKR---TADKDGQMIRSSAIEQGVPLFTALDTAKAMLTVL 1014
             TLV   R G VQA+VNT+G  R   T   DG  IR  AI +G+PLFT+LDT  A+L V+
Sbjct: 995  GTLVEDIRQGRVQAVVNTMGNTRASLTTATDGFRIRQEAISRGIPLFTSLDTVAAILKVM 1054

Query: 1015 ESRCFNIEAI 1024
            +SR F  + I
Sbjct: 1055 QSRSFTTKNI 1064
 Identities = 141/389 (36%), Positives = 222/389 (56%), Gaps = 16/389 (4%)

Query: 518 ESILVLGSGPIRIGQGVEFDYATVHSVKAIQKAGYEAIIMNSNPETVSTDFSVSDKLYFE 577
           + I+++GSGPI IGQ  EFDYA   +  A+++ GYE +++NSNP T+ TD  ++D +Y E
Sbjct: 8   KKIMIIGSGPIIIGQAAEFDYAGTEACLALKEEGYEVVLVNSNPATIMTDREIADTVYIE 67

Query: 578 PLTFEDVMNVIDLEQPKGVIVQFGGQTAINLAQALSEAG------VTILGTQVEDLDRAE 631
           P+T E V  ++  E+P  ++   GGQT +N+A  LS+ G      V +LGT++  +D+AE
Sbjct: 68  PITLEFVSKILRKERPDALLPTLGGQTGLNMAMELSKTGILEELNVELLGTKLSAIDQAE 127

Query: 632 DRDLFEKALKELGIPQPQGQTATNEEEALEAAKKIGFPVLVRPSYVLGGRAMEIVENKED 691
           DR+LF++  + +  P      AT  EEA+  A KIG+P++V P++ +GG    I + +E+
Sbjct: 128 DRELFKELCESINEPLCASDIATTVEEAINIADKIGYPIIVGPAFTMGGTGGGICDTEEE 187

Query: 692 LREYIRTAVKASPEHPILVDSYIFG-KECEVDAISD-GKSVLIPGIMEHIERAGVHSGDS 749
           LRE +   +K SP    L++  I G KE E + + D   + ++   ME+ +  GVH+GDS
Sbjct: 188 LREIVANGLKLSPVTQCLIEESIAGYKEIEYEVMRDSADNAIVVCNMENFDPVGVHTGDS 247

Query: 750 MAVYPPQQLSKQIQETIAEYTKRLAIGLNCIGMMNVQFVI--KNEQVYVIEVNPRASRTV 807
           +   P Q LS    + + + +  +   L   G  NVQ  +   + +  VIEVNPR SR+
Sbjct: 248 IVFAPSQTLSDNEYQMLRDASLNIIRALKIEGGCNVQLALDPNSYEYRVIEVNPRVSRSS 307

Query: 808 PFLSKVTGIPMAQIATKLILGQTLKDL--GYEDGLY----PQSPLVHIKAPVFSFTKLAQ 861
              SK TG P+A+++ K+ +G TL ++     +  Y    P    V  K   F F K
Sbjct: 308 ALASKATGYPIAKMSAKIAIGMTLDEIINPVTNKTYAMFEPALDYVVAKIARFPFDKFEN 367

Query: 862 VDSLLGPEMKSTGEVMGSDTSLEKALYKA 890
            D  LG +MK+TGEVM    ++E++L KA
Sbjct: 368 GDRHLGTQMKATGEVMAIGRNIEESLLKA 396
```

An alignment of the GAS and GBS proteins is shown below:

```
Identities = 777/1025 (75%), Positives = 896/1025 (86%), Gaps = 1/1025 (0%)

Query: 35   LSLKEEGYQVVLVNSNPATIMTDKDIADKVYIEPITLEFVTRILRKERPDALLPTLGGQT 94
            L+LKEEGY+V+LVNSNPATIMTDK+IADKVYIEP+TLEFV RI+RKERPDA+LPTLGGQT
Sbjct: 1    LALKEEGYKVILVNSNPATIMTDKEIADKVYIEPLTLEFVNRIIRKERPDAILPTLGGQT 60

Query: 95   GLNMAMALSKNGILEELNVELLGTKLSAIDKAEDRDLFKQLMEELNQPIPESEIVNSVEE 154
            GLNMAMALSK GIL++L +ELLGTKLSAID+AEDRDLFKQLM+EL+QPIPES IV +V+E
Sbjct: 61   GLNMAMALSKAGILDDLEIELLGTKLSAIDQAEDRDLFKQLMQELDQPIPESTIVKTVDE 120

Query: 155  AIQFAEQIGYPLIVRPAFTLGGTGGGMCDNQEQLVDITTKGLKLSPVTQCLIERSIAGFK 214
            A+ FA  IGYP+IVRPAFTLGGTGGG+C ++E+L +IT  GLKLSPVTQCLIERSIAGFK
Sbjct: 121  AVTFARDIGYPVIVRPAFTLGGTGGGICSSEEELCEITENGLKLSPVTQCLIERSIAGFK 180
```

```
Query: 215  EIEYEVMRDAADNALVVCNMENFDPVGIHTGDSIVFAPAQTLSDVENQLLRDASLDIIRA 274
            EIEYEVMRD+ADNALVVCNMENFDPVGIHTGDSIVFAP QTLSD+ENQ+LRDASL IIRA
Sbjct: 181  EIEYEVMRDSADNALVVCNMENFDPVGIHTGDSIVFAPTQTLSDIENQMLRDASLKIIRA 240

Query: 275  LKIEGGCNVQLALDPNSFKYYVIEVNPRVSRSSALASKATGYPIAKLAAKIAVGLTLDEV 334
            LKIEGGCNVQLALDP SFKYYVIEVNPRVSRSSALASKATGYPIAKLAAKIAVGLTLDE+
Sbjct: 241  LKIEGGCNVQLALDPYSFKYYVIEVNPRVSRSSALASKATGYPIAKLAAKIAVGLTLDEM 300

Query: 335  INPITKTTYAMFEPALDYVVAKMPRFPFDKFESGDRKLGTQMKATGEVMAIGRNIEESLL 394
            INPIT TTYAMFEPALDYVVAK+PRFPFDKFE G+R+LGTQMKATGEVMAIGRN+EESLL
Sbjct: 301  INPITGTTYAMFEPALDYVVAKIPRFPFDKFEHGERQLGTQMKATGEVMAIGRNLEESLL 360

Query: 395  KACRSLEIGVDHIKIADLDNVSDDVLLEKIRKAEDDRLFYLAEALRRHYSIEKLASLTSI 454
            KACRSLEIGV H ++  L N+SD+ L+ K+ KA+DDRLFYL+EA+RR YSIE+L SLT I
Sbjct: 361  KACRSLEIGVCHNEMTSLSNISDEELVTKVIKAQDDRLFYLSEAIRRGYSIEELESLTKI 420

Query: 455  DSFFLDKLRVIVELEDLLSKNRLDINILKKVKNKGFSDKAIASLWQINEDQVRNMRKEAG 514
            D FFLDKL  IVE+E  L  +   +  LKK K  GFSD+ IA +WQ +E  +R MR
Sbjct: 421  DLFFLDKLLHIVEIEQELQMHVDHLESLKKAKRYGFSDQKIAEIWQKDESDIRAMRHSHS 480

Query: 515  ILPVYKMVDTCASEFDSATPYFYSTYAVENESLISDKASILVLGSGPIRIGQGVEFDYAT 574
            + PVYKMVDTCA+EFD+ TPYFYSTY +ENES+ S+K SILVLGSGPIRIGQGVEFDYAT
Sbjct: 481  LYPVYKMVDTCAAEFDAKTPYFYSTYELENESVQSNKESILVLGSGPIRIGQGVEFDYAT 540

Query: 575  VHSVKAIRESGFEAIIMNSNPETVSTDFSISDKLYFEPLTFEDVMNVIDLEKPEGVILQF 634
            VHSVKAI+++G+EAIIMNSNPETVSTDFS+SDKLYFEPLTFEDVMNVIDLE+P+GVI+QF
Sbjct: 541  VHSVKAIQKAGYEAIIMNSNPETVSTDFSVSDKLYFEPLTFEDVMNVIDLEQPKGVIVQF 600

Query: 635  GGQTAINLAKDLNKAGVKILGTQLEDLDRAENRKQFEATLQALNIPQPPGFTATTEEEAV 694
            GGQTAINLA+ L++AGV ILGTQ+EDLDRAE+R  FE  L+ L IPQP G TAT EEEA+
Sbjct: 601  GGQTAINLAQALSEAGVTILGTQVEDLDRAEDRDLFEKALKELGIPQPQGQTATNEEEAL 660

Query: 695  NAAQKIGYPVLVRPSYVLGGRAMKIVENEEDLRHYMTTAVKASPDHPVLIDAYLIGKECE 754
             AA+KIG+PVLVRPSYVLGGRAM+IVEN+EDLR Y+ TAVKASP+HP+L+D+Y+ GKECE
Sbjct: 661  EAAKKIGFPVLVRPSYVLGGRAMEIVENKEDLREYIRTAVKASPEHPILVDSYIFGKECE 720

Query: 755  VDAISDGQNILIPGIMEHIERSGVHSGDSMAVYPPQTLSETIIETIVDYTKRLAIGLNCI 814
            VDAISDG+++LIPGIMEHIER+GVHSGDSMAVYPPQ LS+ I ETI +YTKRLAIGLNCI
Sbjct: 721  VDAISDGKSVLIPGIMEHIERAGVHSGDSMAVYPPQQLSKQIQETIAEYTKRLAIGLNCI 780

Query: 815  GMMNIQFVIKDQKVYVIEVNPRASRTLPFLSKVTHIPMAQVATKVILGDKLCNFTYGYDL 874
            GMMN+QFVIK+++VYVIEVNPRASRT+PFLSKVT IPMAQ+ATK+ILG  L +  Y   L
Sbjct: 781  GMMNVQFVIKNEQVYVIEVNPRASRTVPFLSKVTGIPMAQIATKLILGQTLKDLGYEDGL 840

Query: 875  YPASDMVHIKAPVFSFTKLAKVDSLLGPEMKSTGEVMGSDINLQKALYKAFEAAYLHMPD 934
            YP S +VHIKAPVFSFTKLA+VDSLLGPEMKSTGEVMGSD +L+KALYKAFEA   H+ +
Sbjct: 841  YPQSPLVHIKAPVFSFTKLAQVDSLLGPEMKSTGEVMGSDTSLEKALYKAFEANNSHLSE 900

Query: 935  YGNIVFTVDDTDKEEALELAKVYQSIGYRIYATQGTAIYFDANGLETVLVGKLGENDRNH 994
            +G IVFT+ D  K EAL LA+ +++IGY+I ATQGTA YF   GL   LVGK+G+   N
Sbjct: 901  FGQIVFTIADDSKAEALSLARRFKAIGYQIMATQGTAAYFAEQGLSACLVGKIGD-AAND 959

Query: 995  IPDLIKNGKIQAVINTVGQNNIDNHDALIIRRSAIEQGVPLFTSLDTAHAMFKVLESRAF 1054
            IP L+++G +QA++NTVG     + D  +IR SAIEQGVPLFT+LDTA AM  VLESR F
Sbjct: 960  IPTLVRHGHVQAIVNTVGIKRTADKDGQMIRSSAIEQGVPLFTALDTAKAMLTVLESRCF 1019

Query: 1055 TLKVL 1059
             ++ +
Sbjct: 1020 NIEAI 1024
 Identities = 145/387 (37%), Positives = 229/387 (58%), Gaps = 16/387 (4%)

Query: 10  IMVIGSGPIVIGQAAEFDYSGTQACLSLKEEGYQVVLVNSNPATIMTDKDIADKVYIEPI 69
           I+V+GSGPI IGQ  EFDY+   +  ++++ GY+ +++NSNP T+ TD  ++DK+Y EP+
Sbjct: 520 ILVLGSGPIRIGQGVEFDYATVHSVKAIQKAGYEAIIMNSNPETVSTDFSVSDKLYFEPL 579

Query: 70  TLEFVTRILRKERPDALLPTLGGQTGLNMAMALSKNGILEELNVELLGTKLSAIDKAEDR 129
           T E V  ++  E+P  ++   GGQT +N+A ALS+ G      V +LGT++  +D+AEDR
Sbjct: 580 TFEDVMNVIDLEQPKGVIVQFGGQTAINLAQALSEAG------VTILGTQVEDLDRAEDR 633
```

```
Query: 130 DLFKQLMEELNQPIPESEIVNSVEEAIQFAEQIGYPLIVRPAFTLGGTGGGMCDNQEQLV 189
           DLF++ ++EL  P P+ +   + EEA++ A++IG+P++VRP++ LGG    + +N+E L
Sbjct: 634 DLFEKALKELGIPQPQGQTATNEEEALEAAKKIGFPVLVRPSYVLGGRAMEIVENKEDLR 693

Query: 190 DITTKGLKLSPVTQCLIERSIAGFKEIEYEVMRDAADNALVVCNMENFDPVGIHTGDSIV 249
           +     +K SP    L++  I G KE E + + D   + L+   ME+ +  G+H+GDS+
Sbjct: 694 EYIRTAVKASPEHPILVDSYIFG-KECEVDAISD-GKSVLIPGIMEHIERAGVHSGDSMA 751

Query: 250 FAPAQTLSDVENQLLRDASLDIIRALKIEGGCNVQLALDPNSFKYYVIEVNPRVSRSSAL 309
             P Q LS    + + + +  +   L   G  NVQ  +   + + YVIEVNPR SR+
Sbjct: 752 VYPPQQLSKQIQETIAEYTKRLAIGLNCIGMMNVQFVI--KNEQVYVIEVNPRASRTVPF 809

Query: 310 ASKATGYPIAKLAAKIAVGLTLDEVINPITKTTYAMFEPALDYVVAKMPRFPFDKFESGD 369
            SK TG P+A++A K+ +G TL ++        Y    P    V  K P F F K    D
Sbjct: 810 LSKVTGIPMAQIATKLILGQTLKDL--GYEDGLY----PQSPLVHIKAPVFSFTKLAQVD 863

Query: 370 RKLGTQMKATGEVMAIGRNIEESLLKA 396
             LG +MK+TGEVM    ++E++L KA
Sbjct: 864 SLLGPEMKSTGEVMGSDTSLEKALYKA 890
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 659

A DNA sequence (GBSx0699) was identified in *S.agalactiae* <SEQ ID 2027> which encodes the amino acid sequence <SEQ ID 2028>. This protein is predicted to be carbamoyl phosphate synthetase small subunit (carA). Analysis of this protein sequence reveals the following:

```
Possible site: 19
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2401(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:CAB89872 GB:AJ132624 carbamoyl phosphate synthetase small
           subunit [Lactococcus lactis]
 Identities = 242/355 (68%), Positives = 305/355 (85%)

Query: 2   KRLLLLEDGSVFEGEAFGADVETSGEIVFSTGMTGYQESITDQSYNGQIITFTYPLIGNY 61
           KRLL+LEDG++FEGEA GA+++ +GE+VF+TGMTGYQESITDQSYNGQI+TFTYP++GNY
Sbjct: 3   KRLLILEDGTIFEGEALGANLDVTGELVFNTGMTGYQESITDQSYNGQILTFTYPIVGNY 62

Query: 62  GINRDDYESIRPTCKGVVIYEWAEYPSNWRQQMTLDEFLKLKGIPGISGIDTRALTKIIR 121
           G+NRDDYESI PTCK VV++E A  PSNWR QM+ DEFLK K IPGI+G+DTRA+TKI+R
Sbjct: 63  GVNRDDYESIHPTCKAVVVHEAARRPSNWRMQMSFDEFLKSKNIPGITGVDTRAITKIVR 122

Query: 122 KHGTMKACLINEGNSIHEALENLQKSVLLNDQIEQVSTKLAYASPGVGKNIVLVDFGLKH 181
           +HGTMKA L+   + +   +  LQ +VL  +Q+E  ST  AY SP  G+ +V+VDFGLKH
Sbjct: 123 EHGTMKASLVQARDEVDHQMSQLQATVLPTNQVETSSTATAYPSPNTGRKVVVVDFGLKH 182

Query: 182 SILRELSQRQCHITVVPHTTTAQEILNLNPDGVLLSNGPGNPEQLPNALQMIQEIQGKIP 241
           SILRELS+R+C++TVVP+ T+A+EIL + PDGV+L+NGPG+P  +P A++MI+E+QGKIP
Sbjct: 183 SILRELSKRECNLTVVPYNTSAKEILEMEPDGVMLTNGPGDPTDVPEAIEMIKEVQGKIP 242

Query: 242 IFGICMGHQLFAKANGAKTYKMTFGHRGFNHAVRHLQTGQVDFTSQNHGYAVSREDFPEA 301
           IFGIC+GHQLF+ ANGA TYKM FGHRGFNHAVR + TG++DFTSQNHGYAVS E+ PE
Sbjct: 243 IFGICLGHQLFSLANGATTYKMKFGHRGFNHAVREVATGRIDFTSQNHGYAVSSENLPED 302

Query: 302 LFITHEEINDKTVEGVRHKYYPAFSVQFHPDAAPGPHDTSYLFDEFINMIDDFQQ 356
```

```
           L ITH EIND +VEGVRHKY+PAFSVQFHPDAAPGPHD SYLFD+F++++D+F++
Sbjct: 303 LMITHVEINDNSVEGVRHKYFPAFSVQFHPDAAPGPHDASYLFDDFMDLMDNFKK 357
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2029> which encodes the amino acid sequence <SEQ ID 2030>. Analysis of this protein sequence reveals the following:

```
Possible site: 43
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3534(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 265/354 (74%), Positives = 309/354 (86%)

Query: 2   KRLLLLEDGSVFEGEAFGADVETSGEIVFSTGMTGYQESITDQSYNGQIITFTYPLIGNY 61
           KRLL+LEDG++FEGE FGAD++ +GEIVF+TGMTGYQESITDQSYNGQI+TFTYPLIGNY
Sbjct: 3   KRLLILEDGTIFEGEPFGADIDVTGEIVFNTGMTGYQESITDQSYNGQILTFTYPLIGNY 62

Query: 62  GINRDDYESIRPTCKGVVIYEWAEYPSNWRQQMTLDEFLKLKGIPGISGIDTRALTKIIR 121
           GINRDDYESI PTCKGVV+ E +   SNWR+QMTLD FLK+KGIPGISGIDTRALTKIIR
Sbjct: 63  GINRDDYESISPTCKGVVVSEVSRLASNWRKQMTLDAFLKIKGIPGISGIDTRALTKIIR 122

Query: 122 KHGTMKACLINEGNSIHEALENLQKSVLLNDQIEQVSTKLAYASPGVGKNIVLVDFGLKH 181
           +HGTMKA + ++G+SI    + L+ +VL  + IEQVSTK AY +PG+GKNIVLVDFGLKH
Sbjct: 123 QHGTMKATMADDGDSIQHLKDQLRATVLPTNTIEQVSTKTAYPAPGIGKNIVLVDFGLKH 182

Query: 182 SILRELSQRQCHITVVPHTTTAQEILNLNPDGVLLSNGPGNPEQLPNALQMIQEIQGKIP 241
           SILRE S+RQC+ITVVP   TA+E+L LNPDG++LSNGPGNPE LP AL MI+ +QGKIP
Sbjct: 183 SILREFSKRQCNITVVPFNITAEEVLQLNPDGLMLSNGPGNPEDLPEALDMIRGVQGKIP 242

Query: 242 IFGICMGHQLFAKANGAKTYKMTFGHRGFNHAVRHLQTGQVDFTSQNHGYAVSREDFPEA 301
           IFGICMGHQLF+ ANGAKT KMTFGHRGFNHAVR + TG++DFTSQNHGYAV R   P+
Sbjct: 243 IFGICMGHQLFSLANGAKTCKMTFGHRGFNHAVREIATGRIDFTSQNHGYAVERSSLPDT 302

Query: 302 LFITHEEINDKTVEGVRHKYYPAFSVQFHPDAAPGPHDTSYLFDEFINMIDDFQ 355
           L +THE+INDKTVEGV+H+ +PAFSVQFHPDAAPGPHD SYLFDEF+ MID ++
Sbjct: 303 LMVTHEDINDKTVEGVKHRDFPAFSVQFHPDAAPGPHDASYLFDEFLEMIDSWR 356
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 660

A DNA sequence (GBSx0700) was identified in *S.agalactiae* <SEQ ID 2031> which encodes the amino acid sequence <SEQ ID 2032>. This protein is predicted to be aspartate carbamoyltransferase (pyrB). Analysis of this protein sequence reveals the following:

```
Possible site: 26
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.3260(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAF72727 GB:AF264709 aspartate transcarbamoylase [Enterococcus
           faecalis]
 Identities = 197/303 (65%), Positives = 250/303 (82%)
```

```
Query: 5   TQTLSLEHFVSLEELSNQEVMSLIKRSIEVKENPSNIGFDKDYYVSNLFFENSTRTHKSF 64
           ++ +SL+H ++ E L+++EVM LI+R+ E K+        ++ Y+ +NLFFENSTRTHKSF
Sbjct: 5   SERISLKHLLTAEALTDREVMGLIRRAGEFKQGAKWHPEERQYFATNLFFENSTRTHKSF 64

Query: 65  EMAELKLGLKTIEFNADTSSVNKGETLYDTILTMSALGLDVCVIRHPDIDYYKELIASPN 124
           E+AE KLGL+ IEF A  SSV KGETLYDT+LTMSA+G+DV VIRH   +YY ELI S
Sbjct: 65  EVAEKKLGLEVIEFEASRSSVQKGETLYDTVLTMSAIGVDVAVIRHGKENYYDELIQSKT 124

Query: 125 IHSAIVNGGDGSGQHPSQSLLDLVTIYEEFGYFKGLKIAIVGDLTHSRVAKSNMQVLKRL 184
           I  +I+NGGDGSGQHP+Q LLDL+TIYEEFG F+GLK+AIVGD+THSRVAKSNMQ+L RL
Sbjct: 125 IQCSIINGGDGSGQHPTQCLLDLMTIYEEFGGFEGLKVAIVGDITHSRVAKSNMQLLNRL 184

Query: 185 GAEIFFSGPKEWYSSQFDEYGQYLPIDQLVDQIDVLMLLRVQHERHDGKGVFSKESYHQQ 244
           GAEI+FSGP+EWY  QFD YGQY+P+D++V+++DV+MLLRVQHERHDGK  FSKE YH +
Sbjct: 185 GAEIYFSGPEEWYDHQFDVYGQYVPLDEIVEKVDVMMLLRVQHERHDGKESFSKEGYHLE 244

Query: 245 FGLTKERYKHLRDTAIIMHPAPVNRDVEIASDLVEADKARIVKQMSNGVYARIAILEAVL 304
           +GLT ER   L+  AIIMHPAPVNRDVE+A +LVE+ ++RIV QMSNGV+ R+AILEA+L
Sbjct: 245 YGLTNERATRLQKHAIIMHPAPVNRDVELADELVESLQSRIVAQMSNGVFMRMAILEAIL 304

Query: 305 NSR 307
           + +
Sbjct: 305 HGK 307
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2033> which encodes the amino acid sequence <SEQ ID 2034>. Analysis of this protein sequence reveals the following:

```
Possible site: 38
>>> Seems to have a cleavable N-term signal seq.

----- Final Results -----
                bacterial outside --- Certainty=0.3000(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty=0.0000(Not Clear) < succ>
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 208/300 (69%), Positives = 249/300 (82%)

Query: 8   LSLEHFVSLEELSNQEVMSLIKRSIEVKENPSNIGFDKDYYVSNLFFENSTRTHKSFEMA 67
           ++L + VS+E L+ +EV+ LI R  E K     I   +   V+NLFFENSTRTHKSFE+A
Sbjct: 26  VALTNLVSMEALTTEEVLGLINRGSEYKAGKVVISDHQKDLVANLFFENSTRTHKSFEVA 85

Query: 68  ELKLGLKTIEFNADTSSVNKGETLYDTILTMSALGLDVCVIRHPDIDYYKELIASPNIHS 127
           E KLGL  ++FNAD S+VNKGE+LYDT+LTMSALG D+CVIRHP+ DYYKEL+ SP I +
Sbjct: 86  EKKLGLTVLDFNADASAVNKGESLYDTVLTMSALGTDICVIRHPEDDYYKELVESPTITA 145

Query: 128 AIVNGGDGSGQHPSQSLLDLVTIYEEFGYFKGLKIAIVGDLTHSRVAKSNMQVLKRLGAE 187
           +IVNGGDGSGQHPSQ LLDL+TIYEEFG F+GLKIAI GDLTHSRVAKSNMQ+LKRLGAE
Sbjct: 146 SIVNGGDGSGQHPSQCLLDLLTIYEEFGRFEGLKIAIAGDLTHSRVAKSNMQILKRLGAE 205

Query: 188 IFFSGPKEWYSSQFDEYGQYLPIDQLVDQIDVLMLLRVQHERHDGKGVFSKESYHQQFGL 247
           ++F GP+EWYS  F+ YG Y+ IDQ++ ++DVLMLLRVQHERHDG   FSKE YHQ FGL
Sbjct: 206 LYFYGPEEWYSEAFNAYGTYIAIDQIIKELDVLMLLRVQHERHDGHQSFSKEGYHQAFGL 265

Query: 248 TKERYKHLRDTAIIMHPAPVNRDVEIASDLVEADKARIVKQMSNGVYARIAILEAVLNSR 307
           T+ERY+ L+D+AIIMHPAPVNRDVEIA  LVEA KARIV QM+NGV+ R+AI+EA+LN R
Sbjct: 266 TQERYQQLKDSAIIMHPAPVNRDVEIADSLVEAPKARIVSQMANGVFVRMAIIEAILNGR 325
```

Based on this analysis, it was predicted that these proteins and their epitopes could be useful antigens for vaccines or diagnostics.

Example 661

A DNA sequence (GBSx0701) was identified in *S.agalactiae* <SEQ ID 2035> which encodes the amino acid sequence <SEQ ID 2036>. Analysis of this protein sequence reveals the following:

```
Possible site: 30
>>> Seems to have no N-terminal signal sequence

----- Final Results -----
              bacterial cytoplasm --- Certainty=0.2392(Affirmative) < succ>
               bacterial membrane --- Certainty=0.0000(Not Clear) < succ>
                bacterial outside --- Certainty=0.0000(Not Clear) < succ>
```

The protein has homology with the following sequences in the GENPEPT database:

```
>GP:AAC06948 GB:AE000708 dihydroorotase [Aquifex aeolicus]
 Identities = 176/422 (41%), Positives = 255/422 (59%), Gaps = 8/422 (1%)

Query: 11  IIKNGLIIDPQSGFNQVSDMLIDQGKIKQISKEIDIKGIPIIDASNKIVAPGLVDIHVHF 70
            I+KNG +IDP        D+L++ GKIK+I K I +    IIDA   IV PG +DIHVH
Sbjct: 5   IVKNGYVIDPSQNLEGEFDILVENGKIKKIDKNILVPEAEIIDAKGLIVCPGFIDIHVHL 64

Query: 71  REPGQTHKENIHTGALSAAVGGFTTVLMMANTNPTISSPEIVKQVKESAAKEAI-KIETV 129
            R+PGQT+KE+I +G+  A  GGFTT++ M NTNP I +  +V  + + +    + ++
Sbjct: 65  RDPGQTYKEDIESGSRCAVAGGFTTIVCMPNTNPPIDNTTVVNYILQKSKSVGLCRVLPT 124

Query: 130 ATITKSLNGKDLVNFEELLEAGVAGFSDDGIPLTDTKVLQEAMNLARKHDVVLSLHEEDP 189
             TITK   GK++ +F  L EAG   F+DDG P+ D+ V+++A+ LA +  V +  H ED
Sbjct: 125 GTITKGRKGKEIADFYSLKEAGCVAFTDDGSPVMDSSVMRKALELASQLGVPIMDHCEDD 184

Query: 190 SLN-GVLGINEHIAQKIYHVCGASGLAEYSMIARDAMIAYQTQAKVHIQHLSSSESVEVV 248
             L  GV  INE     +  +   +  AE   IARD ++A +T   VHIQH+S+  S+E++
Sbjct: 185 KLAYGV--INEGEVSALLGLSSRAPEAEEIQIARDGILAQRTGGHVHIQHVSTKLSLEII 242

Query: 249 DFAQKLGANLTAEVTPQHFSKTENLLLTKGANAKLNPPLRLEKDRQALIDGLKSGVISII 308
            +F ++ G  +T EV P H   TE  +L  GANA++NPPLR ++DR ALI+G+K G+I
Sbjct: 243 EFFKEKGVKITCEVNPNHLLFTEREVLNSGANARVNPPLRKKEDRLALIEGVKRGIIDCF 302

Query: 309 ASDHAPHHIMEKAADNISQAPSGMTGLETSLALGITYLVSTKELSMIDFLAKMTCNPAQL 368
            A+DHAPH   EK  + +  A  G+ GL+T+L   +  L     +S+   +   T NPA++
Sbjct: 303 ATDHAPHQTFEK--ELVEFAMPGIIGLQTALPSALE-LYRKGIISLKKLIEMFTINPARI 359

Query: 369 YGFDAGYLREGGPADIVIFDQAEERIIKAEF-ASKSSNSPFIGDKLKGVIHYTICNGEIV 427
             G D G L+ G PADI IFD  +E I+  E   SKS N+P  G  LKG + YTI +G++V
Sbjct: 360 IGVDLGTLKLGSPADITIFDPNKEWILNEETNLSKSRNTPLWGKVLKGKVIYTIKDGKMV 419

Query: 428 YQ 429
            Y+
Sbjct: 420 YK 421
```

A related DNA sequence was identified in *S.pyogenes* <SEQ ID 2037> which encodes the amino acid sequence <SEQ ID 2038>. Analysis of this protein sequence reveals the following:

```
Possible site: 35

>>> Seems to have no N-terminal signal sequence
     INTEGRAL    Likelihood = -0.80   Transmembrane    76 -   92 (  76 -   92)
     INTEGRAL    Likelihood = -0.00   Transmembrane   286 -  302 ( 286 -  302)

----- Final Results -----
               bacterial membrane --- Certainty= 0.132(Affirmative) < succ>
                bacterial outside --- Certainty= 0.000(Not Clear) < succ>
              bacterial cytoplasm --- Certainty= 0.000(Not Clear) < succ>
```

The protein has homology with the following sequences in the databases:

```
!GB:AE000708 dihydroorotase [Aquifex aeolicus]      316  3e-85

>GP:AAC06948 GB:AE000708 dihydroorotase [Aquifex aeolicus]

 Score =  316 bits (801), Expect = 3e-85
 Identities = 177/422 (41%), Positives = 254/422 (59%), Gaps = 8/422 (1%)

Query: 2   ILIKNGRVMDPKSQRDQVADVLIDGKQIVKIASAIECQEAQVIDASGLIVAPGLVDIHVH 61
            +++KNG V+DP    +   D+L++  +I KI   I   EA++IDA GLIV PG +DIHVH
Sbjct: 4   LIVKNGYVIDPSQNLEGEFDILVENGKIKKIDKNILVPEAEIIDAKGLIVCPGFIDIHVH 63

Query: 62  FREPGQTHKEDIHTGALAAAAGGVTTVVMMANTNPVISDVETLQEVLASAAKEKI-HIYT 120
             R+PGQT+KEDI +G+  A AGG TT+V M NTNP I +   +  +L  +    +  +
Sbjct: 64  LRDPGQTYKEDIESGSRCAVAGGFTTIVCMPNTNPPIDNTTVVNYILQKSKSVGLCRVLP 123

Query: 121 NASVTQAFNGKDVTDFKALLEAGAVSFSDDGIPLESSKVLKEAFDLANANQTFISLHEED 180
              ++T+   GK++ DF +L EAG V+F+DDG P+  S V+++A +LA+     I  H ED
Sbjct: 124 TGTITKGRKGKEIADFYSLKEAGCVAFTDDGSPVMDSSVMRKALELASQLGVPIMDHCED 183

Query: 181 PQL-NGVLGFNEGIAEEHFHFCGATGVAEYSMIARDVMIAYDRQAHVHIQHLSKAESVQV 239
             +L  GV+  NEG              AE   IARD ++A     HVHIQH+S   S+++
Sbjct: 184 DKLAYGVI--NEGEVSALLGLSSRAPEAEEIQIARDGILAQRTGGHVHIQHVSTKLSLEI 241

Query: 240 VAFAQQLGAKVTAEVSPQHFSTTEDLLLIAGTSAKMNPPLRTQRDRLAVIEGLKSGVITV 299
            + F ++ G K+T EV+P H    TE  +L +G +A++NPPLR + DRLA+IEG+K G+I
Sbjct: 242 IEFFKEKGVKITCEVNPNHLLFTEREVLNSGANARVNPPLRKKEDRLALIEGVKRGIIDC 301

Query: 300 IATDHAPHHKDEKTVDDMTKAPSGMTGLETSLSLGLTHLVEPGHLTLMSLLEKMTLNPAL 359
             ATDHAPH   EK + +   A  G+ GL+T+L   L  L   G ++L  L+E  T+NPA
Sbjct: 302 FATDHAPHQTFEKELVEF--AMPGIIGLQTALPSAL-ELYRKGIISLKKLIEMFTINPAR 358

Query: 360 LYGFDAGYLAENGPADLVIFADKQERLITENF-ASKASNSPFIGNKLKGVVKYTIADGEV 418
            + G D G L    PAD+ IF   +E ++ E    SK+ N+P  G  LKG V YTI DG++
Sbjct: 359 IIGVDLGTLKLGSPADITIFDPNKEWILNEETNLSKSRNTPLWGKVLKGKVIYTIKDGKM 418

Query: 419 VY 420
            VY
Sbjct: 419 VY 420
```

An alignment of the GAS and GBS proteins is shown below:

```
 Identities = 269/420 (64%), Positives = 338/420 (80%)

Query: 9   MYIIKNGLIIDPQSGFNQVSDMLIDQGKIKQISKEIDIKGIPIIDASNKIVAPGLVDIHV 68
            M +IKNG ++DP+S  +QV+D+LID  +I +I+  I+ +   +IDAS  IVAPGLVDIHV
Sbjct: 1   MILIKNGRVMDPKSQRDQVADVLIDGKQIVKIASAIECQEAQVIDASGLIVAPGLVDIHV 60

Query: 69  HFREPGQTHKENIHTGALSAAVGGFTTVLMMANTNPTISSPEIVKQVKESAAKEAIKIET 128
            HFREPGQTHKE+IHTGAL+AA GG TTV+MMANTNP IS  E +++V  SAAKE I I T
Sbjct: 61  HFREPGQTHKEDIHTGALAAAAGGVTTVVMMANTNPVISDVETLQEVLASAAKEKIHIYT 120

Query: 129 VATITKSLNGKDLVNFEELLEAGVAGFSDDGIPLTDTKVLQEAMNLARKHDVVLSLHEED 188
             A++T++ NGKD+ +F+ LLEAG   FSDDGIPL  +KVL+EA +LA  +   +SLHEED
Sbjct: 121 NASVTQAFNGKDVTDFKALLEAGAVSFSDDGIPLESSKVLKEAFDLANANQTFISLHEED 180

Query: 189 PSLNGVLGINEHIAQKIYHVCGASGLAEYSMIARDAMIAYQTQAKVHIQHLSSSESVEVV 248
            P LNGVLG NE IA++ +H CGA+G+AEYSMIARD MIAY  QA VHIQHLS +ESV+VV
Sbjct: 181 PQLNGVLGFNEGIAEEHFHFCGATGVAEYSMIARDVMIAYDRQAHVHIQHLSKAESVQVV 240

Query: 249 DFAQKLGANLTAEVTPQHFSKTENLLLTKGANAKLNPPLRLEKDRQALIDGLKSGVISII 308
             FAQ+LGA +TAEV+PQHFS TE+LLL  G +AK+NPPLR ++DR A+I+GLKSGVI++I
Sbjct: 241 AFAQQLGAKVTAEVSPQHFSTTEDLLLIAGTSAKMNPPLRTQRDRLAVIEGLKSGVITVI 300

Query: 309 ASDHAPHHIMEKAADNISQAPSGMTGLETSLALGITYLVSTKELSMIDFLAKMTCNPAQL 368
            A+DHAPHH  EK  D++++APSGMTGLETSL+LG+T+LV    L+++  L KMT NPA L
Sbjct: 301 ATDHAPHHKDEKTVDDMTKAPSGMTGLETSLSLGLTHLVEPGHLTLMSLLEKMTLNPALL 360

Query: 369 YGFDAGYLREGGPADIVIFDQAEERIIKAEFASKSSNSPFIGDKLKGVIHYTICNGEIVY 428
```